

## ABSTRACT

### THE DEVELOPMENT AND VALIDATION OF AN F SCALE FOR AN OBJECTIVE TEST BATTERY ON MOTIVATION

by Roger Clay Thweatt

This investigation was concerned with the development and validation of an F scale for an objective test battery of motivation. The study sampled the 4200 administered test protocols of Michigan eleventh grade students who participated in Farquhar's motivational project. Instrumentation consisted of the Generalized Situational Choice Inventory, the Preferred Job Characteristics Scale, the Human Trait Inventory and the Word Rating List.

Rarity responses (based upon a ten percent or less criterion for item selection) were determined separately for a validation and cross-validation statistically defined total sample of 264 males and 264 females. Items comprising the F scale were based upon commonly selected rarity responses between the above samples. Male and female F scales consisted of twenty-five and seventy-three items, respectively. Male and female F scale reliability coefficients of .729 and .746 were obtained by using Hoyt's analysis of variance technique of estimating internal consistency reliability.

The critical F score for both sexes was determined by plotting respective F distribution curves for misclassified and properly classified over- and under-achievers. The point of overlap where misclassified under-achievers scored as properly classified under-achievers on F items was identified as three rarity responses for males and six for females after cross validation.

To obtain evidence of F scale validity three approaches were examined for the effect of F on: 1) expectancy of response fake; 2) test reliability; and 3) test validity.

Under-achievers selected significantly more F items than over-achievers in both male and female samples. Consequently, the rational of high fake expectancy was clearly substantiated.

The respective critical F scores were applied to a sample of males and females. Individuals possessing F scores as large as or greater than the critical score were excluded from the sample. Hoyt's analysis of variance technique for estimating internal consistency was used to obtain a reliability statement of the GSCI discriminating items before and after application of the F scale. It was hypothesized that further evidence of the effectiveness of the F scale could be determined by its ability to remove unstable individuals who tend to lower instrument reliability by erratic test performance. Theoretically, reliability should increase with exclusion of unreliable subjects. However, the effect of homogeneity of test performance may operate also to reduce reliability. The question was raised as to which has the greater effect on reliability: erratic test performance or homogeneity of test performance. To test the effects of the above question a random sample of subjects equal in magnitude to those identified by F as high fake potential were excluded. The assumption was made that the internal consistency reliability coefficient reduced by random selection should be greater than the reliability coefficient reduced by F selection. However, no significant differences between reliability coefficients were found.

The effects on validity between GSCI scores and standardized grade point averages before and after application of the F scale were determined. Before application of F the male and female validity coefficients were .582 and .243, respectively. After the use of F the

male correlation decreased to .501 and the female validity coefficient increased to .394. No significant differences in correlations were obtained. All correlations, however, were significant from zero at the 3% or better level of confidence.

Linear regression lines were plotted for each sex using GSCI scores and standardized grade point averages to locate placement of high F score individuals among the sample. Eighteen percent of the males and thirty-eight percent of the females selecting high numbers of rarity items fell one standard error of estimate below and above the regression line. Eighty-two percent and forty-one percent of the respective males and females fell in the lower left quadrant of the plot. This area represented location of low achieving, low ability students.

Conclusions of the study were:

1. Under-achieving students select significantly more F items than over-achieving students.
2. Further investigation with the F scale should be conducted before employment of the scale in test battery interpretation occurs, particularly for males.
3. The F scale represents a measure of social conformity.
4. The F scale possesses the ability to tap an academic masculinity-femininity continuum.
5. Re-evaluation of F scale concept and utility in clinical instruments should be conducted.

Copyright by  
ROGER CLAY THWEATT  
1961



THE DEVELOPMENT AND VALIDATION OF AN F SCALE  
FOR AN OBJECTIVE TEST BATTERY ON MOTIVATION

By

Roger Clay Thweatt

A THESIS

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

College of Education

1961

## TABLE OF CONTENTS

CHAPTER	Page
I. THE PROBLEM . . . . .	1
Statement of the Problem . . . . .	8
Delimitations . . . . .	8
Statement of the Hypotheses . . . . .	8
Background of Theory and Research . . . . .	9
Summary Statement of Organization . . . . .	19
II. PREVIOUS ATTEMPTS TO DEVELOP VALIDITY	
KEYS: A REVIEW OF THE LITERATURE . . . . .	20
The L Scale . . . . .	20
The F Scale . . . . .	23
The K Scale . . . . .	28
Subtle-Obvious Keys (S-O) . . . . .	41
Set T Scale . . . . .	44
The B Scale . . . . .	47
Miscellaneous Scales . . . . .	49
Similarity Between Various Scales . . . . .	52
Summary . . . . .	57
III. PROCEDURES . . . . .	58
-Background of Farquhar's Study . . . . .	58
General Design of the Motivational Study . . . . .	59
Instrumentation . . . . .	61
Procedures for the Present Investigation . . . . .	62
Rational of High Fake Expectancy . . . . .	66
Summary . . . . .	67
IV. RESULTS OF THE INVESTIGATION . . . . .	68
Selection of the F Scale . . . . .	68
Sex Differences in F Item Selection . . . . .	69
Distribution of F Items . . . . .	70
Reliability of the F Scale . . . . .	71
Validation of the F Scale . . . . .	72
Summary . . . . .	83
V. SUMMARY AND CONCLUSIONS . . . . .	85
The Problem . . . . .	85
Methodology . . . . .	85
Conclusions . . . . .	87
Implications for Further Research . . . . .	89
BIBLIOGRAPHY . . . . .	91
APPENDIX . . . . .	100

## LIST OF TABLES

TABLE	Page
I. • Intercorrelations of K with Other MMPI Variables . .	37
II. Correlations of K Scale with Other Variables Thought to Be Loaded with the K-Factor . . . . .	37
III. Intercorrelations of Five Scales Thought to Be Loaded with The Test-Taking Attitude, No Item Overlap, n = 150 Normal Males . . . . .	38
IV. Intercorrelations of Four MMPI Validity Scale Indicators and Hy for Normals . . . . .	53
V. Summary of Theory of Need-Achievement and Non- Need-Achievement Motivation Basic to Current Research . . . . .	58
VI. Hypothesized Personality Factors Associated with Academic Achievement. . . . .	59
VII. Number of Rarity Items in Each Test . . . . .	69
VIII. Sex Differences in F Item Selection . . . . .	69
IX. Comparisons of F Item Means, Mean Squares and Sample Number Between Male and Female Misclassi- fied and Properly Classified Groups. . . . .	75
X. Comparisons of T-Values and Significant Levels of One Tailed Tests Between Male and Female Mis- classified and Properly Classified Groups . . . . .	76
XI. Effects on GSCI Internal Consistency Reliability Before and After Application of the Male and Female F Scale . . . . .	78
XII. Effects on the Validity Coefficient Between GSCI Raw Scores and Standardized Grade Point Averages After Application of the Male and Female F Scale. . . . .	79
XIII. Significance of Difference Between Validity Corre- lation Coefficients Before and After Application of the Male and Female F Scale . . . . .	80

## LIST OF FIGURES

FIGURE	Page
I. Methodological Selection of Individuals with Stable Measured Aptitude. . . . .	60
II. Method of Selecting Under- and Over-Achievers. . .	61
III. Theorized Model of Selection of Misclassified Under-Achievers Differentiated by the GSCI. . . . .	64
IV. Selection Procedure for Determining F Item Overlap Score . . . . .	65
V. Cross Validation F Item Frequency Distribution for Males (N = 132). . . . .	70
VI. Cross Validation F Item Frequency Distribution for Females (N = 132) . . . . .	71
VII. GSCI Overlap Points for Over- and Under-Achievers for Each Sex. . . . .	73
VIII. Male Regression Plot . . . . .	81
IX. Female Regression Plot . . . . .	82

## CHAPTER I

### THE PROBLEM

The most important failing of almost all structured objective tests is their susceptibility to faking or lying. In addition, objective tests possess an even greater susceptibility to unconscious self deception and role playing.<sup>1</sup> The possibility of such factors having an invalidating effect upon scores has been noted by many writers.<sup>2-26</sup> One of the assumed advantages of the projective methods is that they are relatively

---

<sup>1</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI, " Journal of Applied Psychology, 1946, 30:525-564.

<sup>2</sup>C. R. Adams, "A New Measure of Personality, " Journal of Applied Psychology, 1941, 25:141-151.

<sup>3</sup>G. W. Allport, "A Test for Ascendancy-Submission, " Journal of Abnormal Psychology, 1928, 23:118-136.

<sup>4</sup>G. W. Allport, "The Use of Personal Documents in Psychological Science, " Social Science Research Council Bulletin, 1942, Number 42.

<sup>5</sup>R. G. Bernreuter, "Validity of the Personality Inventory, " Personality Journal, 1933, 11:383-386.

<sup>6</sup>R. G. Bernreuter, "Theory and Construction of the Personality Inventory, " Journal of Social Psychology, 1933, 4:387-405.

<sup>7</sup>R. G. Bernreuter, "The Present Status of Personality Trait Tests, " Educational Research Supplement, 1940, 21:160-171.

<sup>8</sup>Marion Bills, "Selection of Casualty and Life Insurance Agents, " Journal of Applied Psychology, 1941, 25:6-10.

<sup>9</sup>E. S. Bordin, "A Theory of Vocational Interests as Dynamic Phenomena, " Educational and Psychological Measurement, 1943, 3:49-65.

<sup>10</sup>P. Eisenberg and A. Wesman, "A Consistency in Responses and Logical Interpretation of Psychoneurotic Inventory Items, " Journal of Educational Psychology, 1941, 32:321-338.

less influenced by such distorting factors. However, even on projectives malingerers perform in a discriminable fashion. (Here the

---

<sup>11</sup>J. P. Guilford and R. B. Guilford, "Personality Factors S, E, and M and Their Measurement, " Journal of Psychology, 1936, 2:109-127.

<sup>12</sup>D. G. Humm and K. A. Humm, "Validity of the Humm-Wadsworth Temperament Scale: With Consideration of the Effects of Subjects' Response-Bias, " Journal of Psychology, 1944, 18:55-64.

<sup>13</sup>D. G. Humm and G. W. Wadsworth, "The Humm-Wadsworth Temperament Scale, " American Journal of Psychiatry, 1935, 91:163-200.

<sup>14</sup>E. L. Kelley, C. C. Miles and L. M. Terman, "Ability to Influence One's Score on a Typical Pencil and Paper Test of Personality, " Character and Personality, 1936, 4:206-215.

<sup>15</sup>D. A. Laird, "Detecting Abnormal Behavior, " Journal of Abnormal Psychology, 1926, 20:128-141.

<sup>16</sup>C. Landis and S. E. Katz, "The Validity of Certain Questions Which Purport to Measure Neurotic Tendencies, " Journal of Applied Psychology, 1934, 18:343-356.

<sup>17</sup>J. B. Maller, "The Effect of Signing One's Name, " School and Society, 1930, 31:882-884.

<sup>18</sup>W. C. Olson, "The Waiver of Signature in Personal Reports, " Journal of Applied Psychology, 1936, 20:442-450.

<sup>19</sup>Saul Rosenzweig, "A Suggestion for Making Verbal Personality Tests More Valid, " Psychological Review, 1934, 41:400-401.

<sup>20</sup>Saul Rosenzweig, "A Basis for the Improvement of Personality Tests with Special Reference to the M-F Battery, " Journal of Abnormal and Social Psychology, 1938, 33:476-488.

<sup>21</sup>F. L. Ruch, "A Technique for Detecting Attempts to Fake Performance on a Self-Inventory Type Personality Test. " In Quinn McNemar and M. A. Merrill, Studies in Personality (New York: McGraw-Hill, 1942), pp. 229-234.

<sup>22</sup>E. K. Strong, Vocational Interests of Men and Women (Stanford: Stanford University Press, 1943).

<sup>23</sup>P. M. Symonds, Diagnosing Personality and Conduct (New York: Appleton-Century, 1932).

<sup>24</sup>P. E. Vernon, "The Attitude of the Subject in Personality Testing, " Journal of Applied Psychology, 1934, 18:165-177.

clues are extreme cautiousness and hesitancy, rejection of cards, and a minimization of response in general.)<sup>27</sup>

The existence of a distorting influence in test-taking attitude is so obvious that it has hardly been thought necessary to establish it experimentally. However, a number of investigations have empirically demonstrated the effect. Frenkel-Brunswik investigated tendencies to self-deception in rating oneself, finding in some cases marked negative relations between self judgments and the evaluation of others.<sup>28</sup>

Hendrickson reported that a group of teachers earned significantly more stable, dominant, extroverted and self sufficient scores on the Bernreuter scales when instructed to take the test as though they were applying for a position, than when under more neutral instructions.<sup>29</sup> On tests of mental ability, malingerers try more items and make more errors than do intellectually inadequate persons. Malingerers also fail items that handicapped persons pass, and pass items that the defectives fail.<sup>30</sup> A comparative study of malingerers and authentic psychiatric cases using the Cornell Selectee Index and a shortened form of the Shipley

---

<sup>25</sup>J. N. Washburne, "A Test of Social Adjustment," Journal of Applied Psychology, 1935, 19:125-244.

<sup>26</sup>R. R. Willoughby and M. E. Morse, "Spontaneous Reactions to a Personality Inventory," American Journal of Orthopsychiatry, 1936, 6:562-575.

<sup>27</sup>H. G. Gough, "Simulated Patterns on the MMPI," Journal of Abnormal and Social Psychology, 1947, 42:215.

<sup>28</sup>E. Frenkel-Brunswik, "Mechanisms of Self-Deception," Journal of Social Psychology, 1939, 10:409-420.

<sup>29</sup>G. Hendrickson, "Attitudes and Interests of Teachers and Prospective Teachers," (paper read before Section Q, AAAS, Atlantic City, December 27, 1932).

<sup>30</sup>W. A. Hunt and H. J. Older, "Detection of Malingering through Psychometric Tests," United States Naval Medical Bulletin, 1943, 41:1318.

Personality Inventory found that malingerers scored significantly higher on both tests.<sup>31</sup> Ruch showed that college students could fake extroversion on the Bernreuter to the extent of achieving a median at the 98th percentile on Bernreuter's norms, as contrasted with a naive median at the 50th percentile.<sup>32</sup> Bernreuter found that college students could produce marked shifts in their Bernreuter scores in the socially approved direction. He interpreted this finding, however, as indicating the comparative unimportance of the faking tendency. His reasoning was that had the need for giving socially approved responses operated in the first administration to any extent, the effect of special instructions to take this attitude should not have been great.<sup>33</sup> To Meehl and Hathaway this reasoning seemed rather tenuous, inasmuch as the occurrence of a shift merely shows that conscious and permitted faking can produce greater effects than those which may have been operating in the naive original testing.<sup>34</sup> Meehl and Hathaway further state:

The insignificant correlations between naive and faked scores were also used by Bernreuter to support his view, an argument which is not comprehensible in view of the gross skewness of the faked scores. What is clear from his investigation is that people are able to influence their scores to a considerable extent if they choose to, and that the average student's stereotype of what is socially desirable seems to be an individual who is dominant, self sufficient and stable.<sup>35</sup>

---

<sup>31</sup>W. A. Hunt, "The Detection of Malingerers: A Further Study," United States Naval Medical Bulletin, 1946, 46:249.

<sup>32</sup>F. L. Ruch, "A Technique for Detecting Attempts to Fake Performance on a Self-Inventory Type of Personality Test." In Quinn McNemar and M. A. Merrill, Studies in Personality (New York: McGraw-Hill, 1942), pp. 229-234.

<sup>33</sup>R. G. Bernreuter, "Validity of the Personality Inventory," Personality Journal, 1933, 11:383-386.

<sup>34</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:525-526.

<sup>35</sup>Ibid.



Bordin reports that students acquainted with the occupational groupings included in the Strong Vocational Interest Blank were able to simulate certain specified occupational types, even though the students were unfamiliar with the mechanics of scoring.<sup>36</sup> He points out that one factor determining the profile on a test of this kind is the degree of acceptance of an occupational stereotype as a self-description. It is as if a person would ask himself the question, "Who am I?" and then answer the test items in a manner consistent with the resulting self-conception. A second important factor is the degree of knowledge of the true occupational stereotype. This factor will determine the clarity of the obtained interest pattern.<sup>37</sup> Similarly, if a subject is attempting to respond as a psychoneurotic on a personality inventory, the success of the trial will be largely influenced by his understanding of the neurotic syndrome in its intimate as well as its obvious aspects.<sup>38</sup>

Clinical observations substantiate the above conception of malingering. Ossipov states that every malingerer is an actor who portrays an illness "as he understands it." The malingerer goes to extremes, apparently believing that the more eccentric his behavior, the more disordered he will be thought to be. In addition to exaggeration of symptoms, the malingerer tends to act as a "state" or an episode, but not a disease. For this reason Ossipov emphasizes that the entire clinical picture must be carefully evaluated, especially the configuration of symptoms, in distinguishing a feigned from a genuine illness.<sup>39</sup>

---

<sup>36</sup>E. S. Bordin, "A Theory of Vocational Interests as Dynamic Phenomena," Educational and Psychological Measurement, 1943, 3:57.

<sup>37</sup>Ibid., p. 54.

<sup>38</sup>H. G. Gough, "Simulated Patterns on the MMPI," Journal of Abnormal and Social Psychology, 1947, 42:216.

<sup>39</sup>V. P. Ossipov, "Malingering: The Simulation of Psychosis," Bulletin of the Menninger Clinic, 1944, 8:39-42.

Maller,<sup>40</sup> Metfessel,<sup>41</sup> Olson,<sup>42</sup> and Spencer<sup>43</sup> have studied the effects of anonymity on responses to self-rating situations and found that the requirement of signing one's name has a definite effect on scores. Kelly, Miles and Terman demonstrated the great ease with which scores on the Terman-Miles Masculinity-Femininity Test could be faked in either direction once the subjects had been informed concerning the secret of what the test measured.<sup>44</sup> Strong,<sup>45</sup> Bills,<sup>46</sup> Steinmetz,<sup>47</sup> as well as Bordin,<sup>48</sup> have presented evidence of the ability of subjects to distort their interest patterns when taking the Strong Vocational Interest Blank.

There are several reports of MMPI simulation in the literature. Benton had nine homosexuals who were positively identified on the Mf scale retake the test and try to conceal their femininity. Six of the

---

<sup>40</sup>J. B. Maller, "The Effect of Signing One's Name, " School and Society, 1930, 31:882-884.

<sup>41</sup>M. Metfessel, "Personality Factors in Motion Picture Writing, " Journal of Social and Abnormal Psychology, 1935, 30:333-347.

<sup>42</sup>W. C. Olson, "The Waiver of Signature in Personal Reports, " Journal of Applied Psychology, 1936, 20:442-450.

<sup>43</sup>D. Spencer, "Frankness of Subjects on Personality Measures, " Journal of Educational Psychology, 1938, 29:26-35.

<sup>44</sup>E. L. Kelley, C. C. Miles and L. M. Terman, "Ability to Influence One's Score on a Typical Pencil and Paper Test of Personality, " Character and Personality, 1936, 4:206-215.

<sup>45</sup>E. K. Strong, Vocational Interests of Men and Women (Stanford: Stanford University Press, 1943).

<sup>46</sup>Marion Bills, "Selection of Casualty and Life Insurance Agents, " Journal of Applied Psychology, 1941, 25:6-10.

<sup>47</sup>H. C. Steinmetz, "Measuring Ability to Fake Occupation Interest, " Journal of Applied Psychology, 1932, 16:123-130.

<sup>48</sup>E. S. Bordin, "A Theory of Vocational Interests as Dynamic Phenomena, " Educational and Psychological Measurement, 1943, 3:49-65.

nine participants were able to bring their Mf scores within normal limits.<sup>49</sup> Meehl and Hathaway had 54 psychological trainees take the MMPI as if trying to avoid being drafted for military service, and obtained F scale T-scores of 78 or higher in 96 percent of the cases. In addition to high F scores, most of the profiles would have been clinically invalidated because of their highly unusual configurations.<sup>50</sup>

Meehl and Hathaway suggest that it is quite possible that in developing personality questionnaires constructed in the traditional, a priori fashion and refined by statistical manipulation, the test-maker is merely pooling sets of items to differentiate among people with respect to various test-attitude continua of little or no psychiatric relevance. The underlying disposition which leads a subject to respond in a certain way to such questions may or may not be identical with the dispositions recognized as clinical variables, nor with those that might be suggested by the item content. To Meehl and Hathaway it is quite clear on present evidence that identification cannot be established by an assumed equivalence between non-test behavior and the verbal report. Hence, both a priori selection of items and the psychological naming of a statistically homogeneous scale from its item content are fraught with possibilities of error.<sup>51</sup>

Guilford has explicitly called attention to the importance of the problem of test-taking attitudes as "factors" when he says:

We must constantly remember that the response of a subject may not represent exactly what the question implies in its most obvious meaning. Subjects respond to a question as at the moment they think they are, with perhaps a lack of

---

<sup>49</sup>A. L. Benton, "The MMPI in Clinical Practice," Journal of Nervous and Mental Disorders, 1945, 102:416-420.

<sup>50</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:550-551.

<sup>51</sup>Ibid., pp. 553-554.

insight in many cases as to their real position on the question. They also respond as they would like themselves to be and as they would like others to think them to be and as they wish the examiner to think them to be. They also respond with some regard to self-consistency among their own answers. Whether these determining factors are sufficiently constant to set up individual differences which are uniform in character and so constitute common factors in themselves is difficult to say. Should any one of them be so pervasive it should introduce an additional vector in the factor analysis.<sup>52</sup>

### Statement of the Problem

The purpose of this investigation was: 1) to determine whether or not an F scale validity key could be developed for an objective test battery on motivation; and 2) to validate the developed key.<sup>53</sup>

### Delimitations

This study used the administered test protocols of eleventh grade Michigan high school students who were participants in Farquhar's motivational research investigation.<sup>54</sup>

### Statement of the Hypotheses

The following hypotheses were examined in this investigation:  
1) an F scale can be developed for each sex for four of the six

---

<sup>52</sup>J. P. Guilford and R. B. Guilford, "Personality Factors S, E, and M and Their Measurement," Journal of Psychology, 1936, 2:118.

<sup>53</sup>W. W. Farquhar, "A Comprehensive Study of the Motivational Factors Underlying Achievement of Eleventh Grade High School Students," (East Lansing: Approved Research Application of the Commissioner of Education, United States Office of Education, November 1, 1959). 15 pp. (Mimeographed.)

<sup>54</sup>Ibid.

inventories comprising the objective test battery on motivation; and 2) the F scale can differentiate protocols of those test-takers who were uncooperative, who could not comprehend the test items, who made clerical errors, and who intentionally placed themselves in a bad light.

### Background of Theory and Research

Among the many authors who recognize the problem of detection of malingering and falsification on objective tests there are but a few who have made specific suggestions for its solution. The inclusion of special exhortations to frankness and objectivity in the test directions themselves is common, but there is no evidence as to its effectiveness.<sup>55</sup> Obviously, if a subject is consciously determined to fake, he will do so; whereas if his motivation to distortion is of a more subtle, non-verbalized nature, such exhortations can hardly be expected to be efficacious. Another method is to attempt to disguise the content of items, so that the significance of a given response is less obvious. Traditional approaches to the measurement of personality render this technique practically impossible, inasmuch as the items are selected to begin with for their obvious psychological significance. Hence, unless the items are changed so greatly as to no longer elicit the desired information, they will almost inevitably continue to betray their origin. An effective use of a set of subtle items is only possible when the initial item pool is large and the initial selection of items is ruthlessly empirical. Those items whose significance would not have been guessed by the test-maker will then be equally mysterious to the testee. The presence of projective and role playing components of test-taking

---

<sup>55</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI, " Journal of Applied Psychology, 1946, 30:527.

behavior should be recognized in objective personality inventories.<sup>56</sup>

A spurious anonymity using secret coding for identifying the testee is a possibility suggested by the studies cited above, but is clinically impractical and the deception involved is not desirable.<sup>57</sup> Lacking anonymity, it has been suggested by Olson that the name be signed at the conclusion of the test administration instead of at the top of the page.<sup>58</sup> This suggestion was carried into practice by Maller in his Character Sketches.<sup>59</sup> In addition, he also stated the questions in the third person (indirect) form, requiring the subject to indicate whether he was the same or different from the person described. Maller presents evidence that this procedure aroused considerable less annoyance in his subjects; however, direct proof that this decrease in annoyance led to increased validity is lacking. Meehl expresses doubt whether or not the removal of personal reference is wholly desirable because there is reason for believing that the same role playing and self-deception which operate to invalidate some of the measurements are an important factor in making other measurements possible.<sup>60</sup>

Another technique for reducing the effect of signing one's name is to have the items printed on cards which are then sorted by the subject. Such a procedure makes all writing unnecessary and it is assumed that

<sup>56</sup>P. E. Meehl, "The Dynamics of Structured Personality Tests," Journal of Clinical Psychology, 1945, 1:296-303.

<sup>57</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:527.

<sup>58</sup>W. C. Olson, "The Waiver of Signature in Personal Reports," Journal of Applied Psychology, 1936, 20:442-450.

<sup>59</sup>J. B. Maller, Character Sketches (New York: Bureau of Publications, Teachers College, Columbia University, 1932).

<sup>60</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:527-528.

the feeling that one is making a permanent record of his personal failings is lessened. This method has been employed by Maller in a revised test (Personality Sketches) and by Hathaway and McKinley in the MMPI.<sup>61, 62</sup> However, evidence supporting the above assumption concerning the increased validity of performance is lacking.

Although all of these strategems may have a considerable value particularly in the aggregate, the fact still remains that they do not by any means remove the possibility of faking. What is much more important, they are mainly directed at the sort of conscious falsehood which most investigators have stressed, while ignoring the more subtle tendencies to self deception which are probably of even greater importance in affecting scores. Also, they neglect to stress the existence of trends in the opposite direction--namely those trends which exaggerate the apparent abnormality or maladjustment of the individual. Meehl and Hathaway state that it is only natural that the tendency of a testee to put himself in a favorable light should have received more attention than the contrary tendency. However, there is considerable evidence that this latter tendency does exist and that it is a much more important factor in determining scores on inventories than has generally been supposed.<sup>63</sup>

It is also probable that certain systematic differences in item-interpretation, not necessarily a function of personality dynamics of the defensive or self-critical sort but relatively neutral psychologically (semantic variation), lead to score deviations which are misleading.

---

<sup>61</sup>J. B. Maller, "Personality Tests." In J. M. Hunt, Personality and the Behavior Disorders (New York: Ronald Press, 1944).

<sup>62</sup>S. R. Hathaway and J. C. McKinley, "A Multiphasic Personality Schedule: I. Construction of the Schedule," Journal of Psychology, 1940, 10:249-254.

<sup>63</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:528-529.

Such problems have been investigated by Benton,<sup>64</sup> Eisenberg,<sup>65</sup> and Eisenberg and Wesman.<sup>66</sup>

A more fruitful attitude was taken by Rosenzweig in which he reiterated the fact of untrustworthiness of self-ratings and indicated that instead of trying to eliminate completely these sources of error, the test-maker should recognize them and attempt to correct for them in interpreting the results. He says:

Astute phraseology in the instructions and questions of the test have sometimes been resorted to, but such expedients are rarely very effective. Might it not be more effective to recognize at the outset that such tests have certain limitations that can never be completely circumvented and then go on to the measurement of these limiting factors themselves, thus obtaining information by which a correction may be applied to the subject's answers?<sup>67</sup>

Rosenzweig's specific proposal for achieving this end was to include among the usual self-rating items a set of items of the form "I should like to be the sort of man who . . . , " on the theory that if the test-maker knew something of the strength of certain "ideal-self" trends in the person, the investigator could make appropriate correction for these trends in interpreting responses to the traditional items. Rosenzweig, however, never carried this idea into practice. On the other hand, Meehl and Hathaway consider that this approach would be relatively ineffective, since they feel what is desired is not a statement of the strength or number of ideals for the self, but a measure of the extent to which they are allowed to distort responses. In other words,

---

<sup>64</sup>A. L. Benton, "The Interpretation of Questionnaire Items in a Personality Inventory, " Archives of Psychology, 1935, Number 190.

<sup>65</sup>P. Eisenberg, "Individual Interpretation of Psychoneurotic Inventory Items, " Journal of Genetic Psychology, 1941, 25:19-40.

<sup>66</sup>P. Eisenberg and A. Wesman, "A Consistency in Responses and Logical Interpretation of Psychoneurotic Inventory Items, " Journal of Educational Psychology, 1941, 32:321-338.

<sup>67</sup>Saul Rosenzweig, "A Suggestion for Making Verbal Personality Tests More Valid, " Psychological Review, 1934, 41:400-401.



a subject might easily have lofty ideals verbally expressed, but might be too honest, insightful, objective or self critical to distort his responses into agreement with these ideals.<sup>68</sup>

Maller attempted to solve this problem in another way in his Character Sketches by including a small set of items which were supposed to measure the subject's "readiness to confide." The occurrence of normal, well-adjusted scores in combinations with a low-measured "readiness to confide" would lead one to be skeptical of the validity of the measurement.<sup>69</sup> However, the "readiness to confide" items were themselves self ratings on readiness. In the later form called Personality Sketches Maller does not make use of the "readiness to confide" concept, so it may be assumed that it was unsuccessful or at least did not materially improve validity.

Meehl and Hathaway, carrying Rosenzweig's thinking to its logical conclusion, consider the obvious procedure to follow is to give the subject a good chance to distort his answers in accordance with some self picture or conscious facade, and observe the extent to which he does so.<sup>70</sup> Of course, the difficulty here is that such a procedure requires a knowledge of the objective and subjective facts which are usually inaccessible. Here there are apparently three possibilities open to the test builder. First, he may sidestep the problem of getting directly at the objective truth, and attempt to establish falsehood by obtaining internal contradictions, a technique employed by Maller in his earlier test. Cady, in his application of a modified form of the Woodworth

---

<sup>68</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:529.

<sup>69</sup>J. B. Maller, Character Sketches (New York: Bureau of Publications, Teachers College, Columbia University, 1932).

<sup>70</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:530.

Psychoneurotic Inventory to the measurement of juvenile incorrigibility, had earlier made use of repeated items to increase reliability of the scores (although the aim of detecting inconsistency of the fake sort was not explicit in his rationale).<sup>71</sup> Each question appeared twice, once in each section of the test, except that in the second appearance the question was phrased in the negative. Theoretically the subject's response should also be reversed, and the number of failures to reverse is an indication of some inconsistency. Hence, a measure of non-cooperation or dishonesty would be obtained. The inconsistency score obtained in this way was to be subtracted from the adjustment score to get a sort of corrected score as proposed by Rosenzweig. However, Meehl and Hathaway point out that it is by no means obvious that the shift to a negative form of item will leave the projective properties of the stimulus simply reversed in meaning; so that the fact of an inconsistency in the strict logical sense would not necessarily imply lack of cooperation or dishonesty. However, it would seem reasonable that a very large number of such inconsistent pairs would cast grave suspicion upon the scores, either for dishonesty or some equally serious reason.<sup>72</sup> This technique also was abandoned by Maller in his revised instrument.

The second method of using distortion is to present opportunities for answering in an extremely favorable way, but in a way which could almost certainly not be true. This idea was employed by Hartshorne and May in the Character Education Inquiry.<sup>73</sup> If it is assumed that there are very few aspects of behavior for which one could have complete confidence that no subject would be "ideal" in them, it is necessary to

---

<sup>71</sup>V. M. Cady, "The Estimation of Juvenile Incorrigibility," Journal of Delinquency Monographs, 1923, Number 2.

<sup>72</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:530.

<sup>73</sup>H. Hartshorne and M. A. May, Studies in Deceit (New York: Macmillan, 1928).

present a considerable number of such opportunities and progressively reduce the probability that any individual would be as described. In this sense, everyone would possess at least a few highly desirable traits, and no one would be the possessor of all. Without knowing anything whatsoever about a particular person, the test-maker could write on rational grounds a list of extremely good and rare human qualities which is statistically absurd to suppose will all or in large part be possessed by the individual. If the testee says, however, that he has all or a great many of them, it can be decided that he is not telling the truth. The answers to these items could yield strong evidence for deception.<sup>74</sup>

The Humm-Wadsworth Temperament Scales has made use of the socially-desirable response method.<sup>75</sup> Humm and Wadsworth deserve credit for having been among the first investigators of structured personality measurement to lay great stress upon the problem of detecting non-cooperation and distortion of response when evaluating a particular profile of scores. They were also among the first to adopt an explicit and uncompromising empiricism in selecting items from a large initial pool. The two scales which serve as "checks" or "correctors" for the remainder of the profile on the Humm-Wadsworth are the "normal" component and the "no-count." The "normal" component attempts to assess the strength of a general inhibiting, controlling or normalizing factor in personality which Humm and Wadsworth considered always present to act as a "brake" upon strong abnormal tendencies on the other variables. This means that in interpreting a given profile the significance of any deviation on one of the abnormal components must be established with the size of the normal score in mind. However, Meehl

---

<sup>74</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:531.

<sup>75</sup>D. G. Humm and G. W. Wadsworth, "The Humm-Wadsworth Temperament Scale," American Journal of Psychiatry, 1935, 92:163-200.

and Hathaway question Humm and Wadsworth's claim for the normal component.<sup>76</sup>

The "no-count" is based upon the number of items to which the subject responds in the negative. Inasmuch as approximately 76 percent of the scored items of the Humm-Wadsworth are "obviously" suggestive of abnormality when replied to affirmatively, the "no-count" is to some extent a measure of the testee's tendency to avoid, consciously or otherwise, saying "bad" things about himself when taking the test. That this relationship occurs is further supported by the tendency for the no-count to correlate positively (.77) with the normal component and negatively with the various abnormal components.<sup>77</sup> If the no-count is excessively great, the inference is that the subject has responded in a very defensive or possibly stereotyped fashion; and therefore the particular testing is of doubtful validity. Humm and Wadsworth state that as high as 25 or 30 percent of normals seem to invalidate their scores in this way, a proportion which seems to Meehl and Hathaway to be impractically high for clinical purposes.<sup>78</sup> Later, Humm, Storment and Iorns attempted to reduce the proportion of useless tests by a "correction" for the no-count based upon multiple regression procedures.<sup>79</sup> A study of hospitalized psychiatric cases by Arnold indicated that even the exclusion of cases with invalid no-count did not result in any greater

---

<sup>76</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:531.

<sup>77</sup>D. G. Humm and G. W. Wadsworth, "The Humm-Wadsworth Temperament Scale," American Journal of Psychiatry, 1935, 92:174.

<sup>78</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:531.

<sup>79</sup>D. G. Humm, R. C. Storment and M. E. Iorns, "Combination Scores for the Humm-Wadsworth Temperament Scale," Journal of Psychology, 1939, 7:227-253.

v

t

r

p

s

a

w

e

A

t

s

te

ti

m

in

in

e

th

—

T

th

V

A

P

M

M

V<sub>2</sub>

validity clinically than was obtained using all cases.<sup>80</sup> Humm stated that improved multiple regression techniques have resulted in a marked reduction in the proportion of test misses and of uninterpretable profiles.<sup>81</sup>

Washburn, in revising his Test of Social Adjustment, included a set of 21 items modeled after the "lie" items of Hartshorne and May and referred to the total score on this set as "objectivity." This score was included to detect both lying and unintentional inaccuracy. An extremely low objectivity score was said to invalidate the test as a whole. A weighted objectivity score was included in the total score on the entire test.<sup>82</sup>

The third technique available is the empirical derivation of a fake scale by making use of the item shifts obtained when persons take a test under normal naive conditions and then are retested with instructions to fake. This method has been used by Ruch to construct an "honesty" key for the Bernreuter.<sup>83</sup> To Meehl and Hathaway it is interesting that such a procedure "so logical and straightforward, invented to solve a problem so obvious and insistent, should have been employed for the first time over twenty years after the appearance of the first personality inventory."<sup>84</sup> Ruch says:

---

<sup>80</sup>D. A. Arnold, "The Clinical Validity of the Humm-Wadsworth Temperament Scale in Psychiatric Diagnosis," (unpublished Doctor's thesis, University of Minnesota, Minneapolis, 1942).

<sup>81</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:532.

<sup>82</sup>J. N. Washburne, "A Test of Social Adjustment," Journal of Applied Psychology, 1935, 19:125-244.

<sup>83</sup>F. L. Ruch, "A Technique for Detecting Attempts to Fake Performance on a Self-Inventory Type of Personality Test." In Quinn McNemar and M. A. Merrill, Studies in Personality (New York: McGraw-Hill, 1942 ), pp. 229-234.

<sup>84</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:532-533.

The argument is rather simple. If answers to items on a test like the Bernreuter can be faked at all, the chances are that some are easier to fake than others. Therefore, it should be possible to give each item a weight to represent the extent to which it can be faked by the average college student. This was done by tabulating the frequency of each answer to each question for the standard condition and for the influenced condition. These frequencies were converted into percentages, and an honesty weight was assigned to each reply according to the magnitude of the critical ratio of the difference between the frequency of the reply in the honest and in the influenced condition.<sup>85</sup>

Ruch seems to have been the first investigator to attempt empirical derivation of a fake key for a question-answer personality inventory.<sup>86</sup>

As was stated earlier by Meehl and Hathaway there is evidence of a tendency on the part of some testees to make themselves appear in a "bad" light in taking personality tests. Such a tendency is difficult to characterize because it may occur on several different bases. A patient in the hospital may engage in a type of malingering for strictly conscious reasons, presenting a profile on a test which shows abnormalities out of all reasonable proportion to what is apparent from other considerations. Again, there may be somewhat general traits of verbal pessimism or self deprecation which act to distort systematically the results of personality measurement. Meehl and Hathaway have dichotomized the test-attitude continuum by the two opposed terms defensiveness and plus-getting. However, they make no implication concerning the degree of conscious, deliberate deception involved in either. The corresponding extremes of deliberate deception are referred to as faking good and faking bad, respectively. It was recognized that, like the defensive tendency, the plus-getting tendency might exist in all degrees from a mild self-criticality or merely objectivity to a deliberate, conscious

---

<sup>85</sup>Ibid., p. 231.

<sup>86</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:533.

attempt to make oneself look abnormal. Whether this represents simply the extreme of a continuum with faking good at the opposite end, or an entirely new and different factor, was undecided by these authors.

Meehl and Hathaway state:

In any case it would be desirable to develop a scale for detecting these test-taking tendencies to put oneself in a bad light when answering a personality inventory, so that allowance might be made in such cases in the light of a deviant score obtained on such a scale.<sup>87</sup>

### Summary Statement of Organization

The remainder of this investigation is presented in four subdivisions: in Chapter II previous attempts to develop validity keys are epitomized; the motivational research study with its instrumentation and the procedures utilized in the present investigation are discussed in Chapter III; in Chapter IV F scale development, sex differences in F item selection, F scale reliability and validation are discussed; and in Chapter V the summary and conclusion of the investigation are presented.

In Chapter II the literature appertaining to validity scale development is reviewed.

---

<sup>87</sup>Ibid., pp. 533-534.



## CHAPTER II

### PREVIOUS ATTEMPTS TO DEVELOP VALIDITY KEYS: A REVIEW OF THE LITERATURE

For the present investigation all of the major validity scales were reviewed. The order of discussion of the various keys was based upon the chronology of their development with the earliest devised scales discussed first. In addition to the major scales, however, several minor scales had to be considered in order to bring about a more lucid understanding of the major keys.

#### The L Scale

The "lie" scale of the MMPI attempts to identify those individuals who try to falsify their score by choosing responses which they feel are most acceptable socially.<sup>1</sup> The original fifteen MMPI items making up the L scale were selected under the inspiration of the work of Hartshorne and May.<sup>2</sup> Each of the items presents a situation desirable socially, but which is rarely true of an individual. It was recognized by Hathaway and McKinley that extremely conscientious persons would frequently have more than the average of the L items validly positive, but it was assumed that for a person to have six or eight such items marked was highly improbable.<sup>3</sup> It was concluded by these investigators

---

<sup>1</sup>S. R. Hathaway and J. C. McKinley, Manual for the MMPI (New York: The Psychological Corporation, 1945).

<sup>2</sup>H. Hartshorne and M. A. May, Studies in Deceit (New York: Macmillan, 1928).

<sup>3</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:537-538.

that the fifteen items of this type scattered among the main body of the items constituted a fairly subtle trap for anyone who wanted to give an unusually good impression of himself.

The standardization procedure revealed that among the various normal groups the mean score on the L items lay between three and five items. The frequency curves were all skewed sharply in the positive direction. Few individuals obtained raw scores of seven or more. Only two or three percent exceeded ten items. These values were arbitrarily called the 60 and 70 T-score points, respectively. As more data were accumulated Hathaway and McKinley concluded that the original tentative assumptions regarding the meaning of L were in the main correct, but other valid interpretations of L in the range from T-score 56-70 also existed. The original arbitrary assignment of T-scores had been too conservative, and more emphasis was placed on the T-score range of 56-60. To Hathaway and McKinley the positive presence of the rise in the L score seemed quite valid as an indicator that the individual taking the test was being dishonest and might be somewhat unreliable. However, if no rise in L was observed, these investigators offered no positive or clear interpretation.<sup>4</sup>

To check the assumption that L would not identify the more sophisticated subject, an experiment was performed with 53 male psychology students. The participants, who had completed a considerable portion of their training in psychology, were asked to take the MMPI twice. The first administration was done in the standard manner. In the second administration the group was asked to make certain that they would be acceptable to army induction. Half the group took it with fake good instructions first, half second. Through this procedure a faked good record and a normal record were both obtained.<sup>5</sup>

---

<sup>4</sup>Ibid.

<sup>5</sup>Ibid.

These records showed no appreciable rise in L. It is also true, however, that the majority of the profiles were only slightly better than the corresponding non-faked profiles.

Although one might conclude from the design of the above experiment that the outcome simply tested the participants' willingness and/or ability to fake good, Hathaway and McKinley held that the results demonstrated that the intent to deceive is not often detectable by L when the subjects are relatively normal and sophisticated.<sup>6</sup>

Cottle, however, found additional support for one of the conclusions of Hathaway and McKinley, namely, that sophisticated, bright individuals tend to score low on L items. In his study with 100 high level college students on the MMPI Cottle found that the mean L score on the card form was 2.54 raw score points and for the booklet form was 2.73 raw score points.<sup>7</sup> However, no conclusions can be reached from this study concerning the hypothesis that the L scale is a valid key in differentiating between individuals who wish to fake good and those who do not.

Hovey discusses three cases of individuals who discovered the scoring purpose of cuts on MMPI cards. He says that the L score in these cases was zero.<sup>8</sup> On the other hand, Cofer and others when studying the effect of malingering on the MMPI found that the instructions to fake a normal profile raised L scores.<sup>9</sup>

From the evidence presented in the literature no definite conclusions can be drawn at this time concerning the efficacy of the L scale. No probability statement regarding its differentiating power is offered.

<sup>6</sup>Ibid.

<sup>7</sup>W. C. Cottle, "Card Versus Booklet Forms of the MMPI, " Journal of Applied Psychology, 1950, 34:255-259.

<sup>8</sup>H. B. Hovey, "Detection of Circumvention in the MMPI, " Journal of Clinical Psychology, 1948, 4:97.

<sup>9</sup>C. N. Cofer, J. Chance and A. J. Judson, "A Study of Malingering on the MMPI, " Journal of Psychology, 1949, 27:491-499.

The literature only rationally suggests that factors of degree of training and psychological sophistication may influence unduly an individual's performance on L items.

### The F Scale

In the original publication of the MMPI the F scale was not presented as an empirically validated variable. Its validity was assumed on a priori grounds. The key was composed of 64 items which were selected because they were answered with relatively low frequency (10 percent or less) in either the true or false direction by the main normal group. The scored direction of response was the one which was rarely made by unselected normals. Additionally, the items were chosen to include a variety of content so that it was unlikely that any particular pattern would cause an individual to answer many of the items in the unusual direction. The relative success of this selection of items, with the deliberate intent of forcing the average number of items answered in an unusual direction downward, was illustrated in the fact that the mean score on the 64 items ran between two and four points for all normal groups. The distribution curve was extremely skewed; the higher F scores approached half the total number of F items. In distributions of normal persons the frequency of scores dropped rapidly at about seven items and was at the two or three percent level of chance by a score of twelve. Because of this quick cutting off of the curve and scores seven and twelve were arbitrarily assigned T-score values of 60 and 70 in the original F table.<sup>10</sup>

From the first Hathaway and McKinley recognized that F represented several interpretations. 1) The subject would need to sort almost all of

---

<sup>10</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI, " Journal of Applied Psychology, 1946, 30:535-536.

the items according to expectation in order for these low scores to result; and any error in recording, such as mistaking true items for false items and the like would raise the F score appreciably. 2) If a subject could not understand what he was reading adequately enough to comprehend fully the answers to these items, the F score would obviously be higher. 3) Persons who were highly individualistic and independent might honestly make infrequent responses to F score items. For example, such individuals might admit to disliking children and not believing their mother was a good woman. 4) It was early discovered that schizoid subjects and subjects who apparently wished to put themselves in a bad light also obtained high scores. Meehl and Hathaway felt that the schizoid group obtained high scores because they said unusual things due to delusional or other aberrant mental states in responding to the items. This was referred to as distortion since it was considered that an impartial study would not justify the patient's placement. Among more normal persons some high scores were also observed where the individual had rather unusual ways of responding to conventional stimuli. For example, to the item, "I have had periods in which I carried on activities without knowing later what I had been doing," most persons answered false. Some individuals, however, included periods of sleep in the implication of the item. One might argue that such ways of thinking are often allied to schizoid mentation generally and that the answers in this case indicate a true abnormality. At the very least, however, the person is responding to some items in a way that differs from that of most individuals. Meehl and Hathaway conclude that such persons might, therefore, not be appropriately approached through this method of personality measurement. To them it seemed reasonable that there are individuals whose habitual ways of reacting to items are so different from other persons that measurement of their personalities through the use of verbal items of this type would reflect the unusualness of their reactions to the items more than any clinical abnormality.<sup>11</sup>

---

<sup>11</sup>Ibid., p. 536.

Clinical experience suggested to Meehl and Hathaway that the usual critical score of T equals 70 was too low in the case of F. They found that scores ranging up to T scores of 80 were often more a reflection of validly unusual symptoms and attitudes than an indication of invalidity in the rest of the profile due to misunderstanding, etc. However, scores above this strongly suggested an invalid record.<sup>12</sup> As a result, it was decided that scores above 70 would indicate the whole record to be invalid, except in the special cases mentioned above of schizoid tendencies. Scores from 60-70 would be considered opened to suspicion; scores from 50-60 would be considered a reliable sign that item comprehension, clerical work, etc., had been satisfactory and that the subject was similar to persons in general.

When the MMPI was administered to incoming servicemen it was possible to consider the F score as evidence of an attempt to malingering and to obtain fallaciously bad scores on the other MMPI scales. To check on this interpretation, a similar study to the investigation conducted on the L scale (see pages 21-22) was devised. A group of 54 service men who had completed a considerable portion of their training in psychology were asked to take the MMPI twice. The group took the MMPI in the standard way and also took it under instructions to assume that they wished to avoid being accepted in the draft; and in order to be rejected they were to obtain adverse scores without giving themselves away. The order was reversed for half of the group for the test administration. Through this plan a faked bad record and a normal record were obtained. The data reveals that 96 percent of the faked bad records had a raw F score of 15 or more (T-score of 78 or greater). The researchers concluded that even these men who were somewhat cognizant of psychological measurements betrayed themselves when they attempted to fake a bad record.<sup>13</sup> The F

---

<sup>12</sup>Ibid., pp. 536-537.

<sup>13</sup>Ibid., p. 537.

scale was a good device for identifying the intentional faking that could be set up in an experimental situation.

Kazan and Sheinberg found, however, that a high F score is rarely an invalidating factor with abnormal subjects. In their investigation of 170 maladjusted, male servicemen, all the items of the F score were not answered in the infrequent direction less than 10 percent of the time by normals, and that the percentage was but little higher for miscellaneous abnormal subjects.<sup>14</sup> Schmidt, likewise, found that the profiles for psychotics rose more sharply at F than for any other clinical group.<sup>15</sup> Schneck also found a high F score less valid in a study of character disorders in an army disciplinary barracks.<sup>16</sup> Cofer and others show that subjects attempting to fake emotional upset were detected easily by the F score.<sup>17</sup> Another study of faking on the MMPI by Gough reports that feigned psychotic curves can be detected by being too low on neurotic scales, too high on psychotic scales, and by a significantly elevated F scale. Gough also found that feigned neurotic curves were identified by high F and low K scores.<sup>18</sup> Mechanical sorting using an index of F minus K correctly selected 82 percent of these feigned neurotic profiles. This study used eleven individuals with a background in psychology or psychiatry as contrasted with controls of thirteen hospitalized paranoid schizophrenics and 57 severe psychoneurotics. Gough concludes that

---

<sup>14</sup>A. T. Kazan and I M. Sheinberg, "Clinical Note on the Significance of the Validity Score F in the MMPI," American Journal of Psychiatry, 1945, 102:181-183.

<sup>15</sup>H. O. Schmidt, "Test Profiles as a Diagnostic Aid: the MMPI," Journal of Applied Psychology, 1945, 29:115-131.

<sup>16</sup>J. M. Schneck, "Clinical Evaluation of the F Scale on the MMPI," American Journal of Psychiatry, 1948, 104:440-442.

<sup>17</sup>C. N. Cofer, J. Chance and A. J. Judson, "A Study of Malinger-ing on the MMPI," Journal of Psychology, 1949, 27:491-499.

<sup>18</sup>H. G. Gough, "Simulated Patterns on the MMPI," Journal of Abnormal and Social Psychology, 1947, 42:215-225.

relatively skilled persons are unable to simulate either a psychoneurotic or psychotic condition on the MMPI in such a way as to avoid detection. Hunt used a group of 109 psychology students and 74 Navy general court martial prisoners to investigate the effect of deliberate deception.<sup>19</sup> He substantiated Gough's discovery that an index of F minus K correctly identifies a substantial proportion of malingered or faked abnormal profiles. However, Hunt concluded that the index was of no use in detection of faked normal profiles.

A later report on the F minus K index by Gough suggests that the sampling distribution of F minus K is reasonably normal and the index is not particularly distorted by psychiatric abnormality.<sup>20</sup> He presents data from Sweetland which support the use of F minus K.<sup>21</sup>

After reviewing the MMPI literature Cottle found that high scores on F are caused by carelessness of the subject, scoring errors, psychotic state, or deliberate faking of an abnormal profile. He, too, concluded that the F scale is of little use in detecting faked normal profiles.<sup>22</sup>

From the evidence presented in the literature Cottle's conclusions concerning the F scale appear rationally sound. However, as in the case of the L scale, no translucent probability statement is offered regarding the differentiating power of F in distinguishing between adequate and inadequate performance (as outlined by Hathaway and McKinley on pages 23-24 above).

---

<sup>19</sup>H. F. Hunt, "The Effect of Deliberate Deception on MMPI Performance," Journal of Consulting Psychology, 1948, 12:396-402.

<sup>20</sup>H. G. Gough, "The F Minus K Dissimulation Index for the MMPI," Journal of Consulting Psychology, 1950, 14:408-413.

<sup>21</sup>A Sweetland, "Hypnotic Neurosis--Hypochondriasis and Depression," Journal of Genetic Psychology, 1948, 39:19-105.

<sup>22</sup>W. C. Cottle, "The MMPI: A Review," Kansas Studies in Education, 1953, 3:6-9.



### The K Scale

Meehl and Hathaway conceptualized two approaches to the problem of identifying the attitude a subject takes toward the items in a personality inventory. First, the investigator may have the subject deliberately assume a generally defined attitude. For example, faking might be directed toward obtaining either adverse or desirable scores. A "normal" set of responses must be obtained relatively simultaneously with the faked responses if a reference point is to be determined. The faked and normal records can then be contrasted for study to discover the items which are most frequently changed from the normal records as contrasted to the fake records. Secondly, the investigator may choose records in which there is presumptive likelihood that a special attitude has been assumed. As in the first approach a "normal" set of responses must be obtained simultaneously for comparison.<sup>23</sup>

Using the direction to fake approach several scales were derived by Meehl and Hathaway. The good and bad fake scales were found to be composed of different sets of items. Either of the procedures provided a scale that would be about as good for the other type of faking as it was for the one from which it was derived when such scales were applied to test cases not used in the original derivations. Using two such scales separately did not materially increase the accuracy of prediction.

In the second line of experimental approach Meehl and Hathaway also derived several subdivisions. Among presumably functional and normal records cases were often identified which were so abnormal as to indicate that the individual should have been hospitalized. The investigators attempted to discover items which would differentiate normal from clinically diagnosed abnormal persons. For the counterpart to this approach they also selected for item analysis hospitalized cases whose

---

<sup>23</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI, " Journal of Applied Psychology, 1946, 30:538-539.

records showed a normal profile. Using this approach Meehl and Hathaway experimentally derived four scales.

Derivation of the L6 Scale. The most important finding of the investigators was that whichever of the methods used, as was with the case with the faked approach above, the resultant scales were about equally effective. These scales were also fairly effective in differentiating the fake group as well. After two years of this experimentation all of the promising scales were cross-validated on a new sample. A single best scale was derived which was originally called L6.<sup>24</sup>

L6 was derived by an item analysis of the responses of 25 males and 25 females in a psychopathic hospital. These subjects' MMPI profiles showed an L scale ("lie" key) of T equal 60 or more. In addition, these individuals were predicted to obtain abnormal profiles because of the clinical diagnoses given to these cases by the psychiatric staff. However, the scores on the MMPI profiles fell within the normal range.

Two restrictions were employed in the selection of the criterion group. All of the individuals were characterized by deviant behavior; however, they obtained relatively normal profiles and were termed "misses" for the MMPI. In addition, all criterion cases were characterized by having a tendency to obtain elevated scores on the L scale.

The item responses of these 50 cases, analyzed separately for males and females, were compared to item frequencies from previous standardization groups. In all, 22 items were chosen as a result of this comparison on the basis of a 30 percent discrepancy between validation and standardization groups.

Because the criterion group was assumed to desire good scores, the larger raw scores on these items were in the same direction as the larger raw scores on the L scale. The item content suggested an attitude of denying worries, inferiority feelings, and psychiatrically unhealthy

---

<sup>24</sup>Ibid., pp. 539-540.

symptoms, together with a disposition to see only good in others as well as oneself.

Cross-validation of L6. Following the final choice of L6 as the best of the scales available, Meehl and Hathaway subjected the validity key to more careful study. Hospital and normal records were examined to discover whether L6 would be helpful in interpreting individual profiles. Relatively few data were found on normal cases, but on hospital cases a fairly extensive symptomatic summary was available. By examining the profile for normalcy it was determined whether or not the L6 deviated in an upward direction. It was assumed that an upward direction indicated that the patient had attempted to place himself in a good light. As a result of this study L6 was judged effective but left much to be desired.<sup>25</sup>

The L6 scale as a measure of defensive and plus-getting attitudes. To the investigators L6 appeared as adequate for the detection of plus-getting as was N (see section below) or any of the other experimental scales. Accordingly, the records of a new series of presumable normal persons showing deviant profiles was examined. The L6 scale again appeared to work at the plus-getting end of the test-attitude continuum. That is to say, a relatively low score on L6 could be used to under-interpret an otherwise deviant profile and thus avoid some of the presumable false positives in the normal population sample. Thus L6 seemed useful at both ends of the test-attitude continuum: defensiveness and plus-getting.

Refinement of the L6 scale. The most outstanding difficulty in the above procedure was that L6 tended to be low on severe depressive or schizophrenic patient records. This led to an under-interpretation in spite of the fact that the patients were grossly abnormal. To correct for the under-interpretation tendency, items were added that would work in the opposite direction. To choose these items Meehl and Hathaway

---

<sup>25</sup>Ibid., pp. 540-542.

studied the item tabulations for the group of psychological trainees above who had attempted to fake good and bad scores. In the above study there were many items which showed no tendency to change with an alteration in the test-taking attitude. The percent of true or false remained constant whether the attitude was the normal one or the faked one. From among these items, a sub-group was chosen which showed differences between schizophrenic and depressive criterion groups and general population normals. Meehl and Hathaway admit that the procedure rested upon the insecure assumption that any item that did not appear to be affected by the test-taking attitude (as approached by a normal person attempting to fake good or bad, but occurred as a frequent item to differentiate depressed or schizophrenic patients) would be useful in correcting the tendency of L6 to go too low for schizophrenic and depressed patients. Such an item was scored in a way that would make it work against the tendency of the L6 scale. Eight items were selected by this method. The effect of adding these eight items to the 22 on L6 was to elevate slightly the mean score of normals and to make it more nearly approach the mean score of abnormal cases on the complex of all 30 items.<sup>26</sup>

Derivation of the K scale. As a final step in the refinement of the L6 scale the above eight items were combined with the 22 L6 items into a single scale which was designated K. The K scale represents the final outcome of many experiments in the field of measuring test attitude.

Meehl and Hathaway state:

The K scale is far from perfect for its purpose as measured by the various available data. Generally speaking it is about as good as any other single scale yet derived. In individual applications it is inferior now to one scale and now to another but the differences are never great enough to be very significant practically and the small number of items in this scale gives it a distinct advantage over one or two of the longer scales such as N.<sup>27</sup>

---

<sup>26</sup>Ibid., p. 543.

<sup>27</sup>Ibid., pp. 543-544.

Because the K scale was derived as a correction scale for improving the discrimination yielded on the already existent MMPI scales, it was not assumed to be measuring anything which in itself was of psychiatric significance.

Meehl and Hathaway considered that it was first necessary to choose criterion cases of the sort on which K could conceivably be of value. It was apparent to these investigators that such cases would be characterized by the presence of what was called borderline profiles, that is, individuals who possessed T-scores between 65-80. In studying hundreds of deviant profiles after the addition of K, almost no individuals were found with T-scores above 80 in the normal sample; and it was not statistically profitable to correct elevations of such magnitude to the point of calling them normal. On the other hand, when a curve showed no elevations at all above 65, even the presence of a high K score did not enable the examiner to form any adequate notion of what the peak would be had the K factor not been operating to distort the results. There were apparently upper and lower limits beyond which deviations on K could not effectively operate. Profiles showing subtest scores above 80 were interpreted as probably abnormal no matter how low K fell. If a profile showed no subtest scores above 65 it was unknown whether a high K meant the profile should be adjusted toward more severe scores or was merely that of an actually normal person who for some reason took a defensive attitude when being tested.

Validation of the K scale. Meehl and Hathaway judged that the kind of curve which gave interpretative difficulty, and which could be improved by knowledge of the influence of K, would be a curve in the doubtful, borderline region. Accordingly, a group of cases from the normal and hospital groups was chosen on the basis of having achieved such borderline curves. For this study all cases in the files showing at least one personality component elevated as high as T equal 65, but with no component elevated to T greater than 80. Among the normals,

there were 71 males and 103 females having such borderline curves. Corresponding to these cases, 129 abnormal males and 208 abnormal females were located with similar borderline profiles. The data for the two sexes were treated separately.

The analysis of these data was in terms of the ability of the K scale, used mechanically, to separate the curves of the actual normals from those of the actual abnormals. The procedure was to arrange the whole set (normals and abnormals combined) for each sex in order of the magnitude of their K scores. The distribution of K was cut on the basis of the proportion of normals and abnormals in the sample, calling all cases above the cut abnormal and all those below normal. Setting up a fourfold table on this basis, a chi-square of 20.436 for the males and 29.540 for the females was obtained. Both of these were highly significant (P equal to less than .0010) with one degree of freedom. If instead of locating an optimal cutting score, the K distribution was cut at the mean of the general population K distribution (T equal 50), the cutting point of the males was unchanged. However, the cutting point for the females shifted enough to lower their chi-square to 17.750, which was still highly significant. If one considered miscellaneous profiles which lie in the borderline range between 65 and 80, regardless of the kind of elevation and irrespective of the clinical diagnosis of those who were clinically abnormal, one could separate them into actual normals and abnormals significantly better than chance by using a cutting score on K. In this instance Meehl and Hathaway emphasized that K was operating chiefly as a suppressor of certain test-taking tendencies, for K by itself did not differentiate unselected normal and abnormal cases. In terms of percentages, it was found that for the males, 72 percent of the abnormals and 61 percent of the actual normals were correctly identified. For the females, 66 percent of the abnormals and 59 percent of the normals were correctly classified. These percentages were based upon the separations

with K equal 50, taking no account of the actual normal-abnormal proportions among the above cases.<sup>28</sup>

Refinement of the K scale. Evidence from examination of the test "misses" disclosed by K in the data, combined with knowledge of the correlation between K and other MMPI scales, indicated that the K correction was more important in some scales than in others. Therefore, it was decided to analyze the borderline groups in terms of the peak elevation of their profiles in the attempt to identify those particular curves on which K could be used with profit.

The entire group of 511 borderline curves (males and females, normals and abnormals pooled) was divided into eight sub-groups, each sub-group composed of cases having the peak score on the same one of the eight MMPI personality components.

The normals and abnormals having borderline curves with the same peak score were then separated mechanically by the use of a cutting score on K. The proportion of cases above the cutting score was determined on the basis of the proportion of actual abnormals versus normals in each sub-group. Meehl and Hathaway state that it was unavoidable in the analysis because the relative proportions of actual normals and abnormals varied widely from scale to scale and the use of the mean of K would have been grossly misleading. For the eight groups studied in this manner, only three showed a significant chi-square (P less than .01). One clinical scale group yielded a chi-square between the 10 percent and 20 percent level of significance. It seemed, therefore, that the K factor could be used with profit in interpreting some kinds of profiles but not others. The failure to discriminate with K when grouping profiles by peak score did not establish that a K correction might not be profitably added to the single scores themselves.<sup>29</sup>

---

<sup>28</sup>Ibid., p. 545.

<sup>29</sup>Ibid., p. 546.

Cross-validation of the K scale. One other validating study was done of K in the original investigation. A group of 22 normals and 22 abnormals who were employed in a previous study was used.<sup>30</sup> The normals consisted of a random selection from a large group of profiles showing any elevation of 70 or over. The abnormals consisted of a heterogeneous group also having at least one subtest score over 70. All cases were chosen randomly from hospital cases. Because these groups had been selected for a different investigation, they had not entered into the derivation of K in any way. Without regard for any other information concerning the profiles, all cases showing K greater than 50 were arbitrarily guessed as abnormals, whereas those with K less than 50 were called normals. The cutting score was also independent of the statistics of the original group. To Meehl and Hathaway the K scale worked phenomenally well here. Of the entire group of 44 cases, 37 were correctly classified when using K in this way, a total of 85 percent "hits." Here it was the purpose to separate normals and abnormals all of whom possessed deviant profiles. The investigators concluded that this percent was quite impressive considering the task set for K. Of the seven errors in classifying, six were false positives (cases of normals showing elevated profiles and K greater than 50 and termed abnormal). The chi-square for the fourfold table of these data was 21.569 which with one degree of freedom was highly significant (P less than .0001). Meehl and Hathaway conclude:

Here we have striking evidence of the validity of K when used to differentiate between deviant curves of actual normals and abnormals. We are not prepared to explain the superiority of this result to that originally encountered, except to say that the range of abnormal scores in the present analysis was from 70-90 whereas in the original analysis borderline scores were defined as lying between 65 and 80. In what way this could make K appear to function more effectively in the one case than the

---

<sup>30</sup>P. E. Meehl, "An Investigation of General Normality Control Factor in Personality Testing," Psychological Monographs, 1945, 59: Number 4.



other is not clear. Also the present study involved only males where K in general seems to work a little better than on females.<sup>31</sup>

Discrepancy in the efficacy of K. The fact that K was less effective when applied to some scales than others would suggest separate interpretations or cutting scores. Furthermore, the classification into normal and abnormal on the basis of a single arbitrary cutting score obviously sacrifices some quantitative information about the actual magnitude of the K score. Meehl and Hathaway, however, did not tend to propose such a cutting method as the most efficient manner of application for K. They simply used that form to indicate that K possessed differentiating power for what it was hoped to differentiate.

With the exception of Hy and D clinical scales the correlations of K with the other MMPI variables were consistently negative. This was to be expected if K represented the defensive, lying, or self-deceptive test-taking attitude it was derived to measure. It might be thought that such low correlations as occur in Table I below would preclude any possibility of the use of K as a suppressor. However, there is a tendency for the scales on which K seems valid by the chi-square test to show higher correlations. But for the use of which K was put, correlations as low as .20 were utilized to yield very significant and useful improvements in discrimination.<sup>32</sup>

Considering the relative unreliability of some of the MMPI variables, the intercorrelations of the K scale with other variables considered loaded with the K factor, are rather impressive: the G scale and plus scale were derived wholly by internal item relationships and without regard to criteria of any non-test behavior; the N scale corrected for self-criticality of certain plus-getters who showed deviant profiles; the Ch scale differentiated hypochondriacs from non-hypochondriacal

---

<sup>31</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI, " Journal of Applied Psychology, 1946, 30:546-547.

<sup>32</sup>Ibid., pp. 548-549.

Table I. Intercorrelations of K with Other MMPI Variables

	Hs	D	Hy	Pd	Pa	Pt	Sc	Ma
Normal males	-.30	.15	.48	-.17	-.07	-.67	-.59	-.36
Normal females	-.35	-.03	.30	-.06	-.02	-.64	-.58	-.28
Abnormal males	-.42	-.29	.11	-.26	-.19	-.60	-.60	-.37
Abnormal females	-.17	-.16	.17	-.21	-.13	-.63	-.58	-.38

abnormals who had elevated H scores; and a sub-set of items (Hy-O) which were chosen because they differentiated a clinical group--hysteria. There was, however, a considerable item overlap among these scales, which tended to raise the correlations. On the other hand, Meehl and Hathaway point out that the K scale was not actually pure for the hypothetical test-taking attitude because it was a composite of the test-taking scale L6 plus eight psychotic items. This would presumably tend to lower the correlations. Accordingly, Meehl and Hathaway substituted L6 for K, removed the item overlap among the scales G, N, Ch, L6 and

Table II. Correlations of K Scale with Other Variables Thought to Be Loaded with the K-Factor

	"+"	G	N	Ch	Hy-O
Normal males	-.64	-.76	-.70	-.67	.81
Normal females	-.62	-.73	-.64	-.63	.78
Abnormal males	-.70	-.75	-.69	-.64	.74
Abnormal females	-.70	-.81	-.72	-.71	.74

Hy-O and calculated correlations among these reduced keys. Table III shows the intercorrelations among these five non-overlapping keys, based upon the responses of 150 unselected normal males between the ages of 26-45, rejecting records of F greater than 80. All scales were scored so as to render the correlations positive.<sup>33</sup>

Table III. Intercorrelations of Five Scales Thought to Be Loaded with The Test-Taking Attitude, No Item Overlap, n = 150 Normal Males

	G	Ch	L6	N
CH	.82			
L6	.76	.71		
N	.78	.73	.66	
Hy-O	.70	.63	.70	.59

This correlation matrix was subjected to a factor analysis and repeated three times in successively approximating the communalities because of the small number of tests. It appeared that one common factor was quite sufficient to account for the intercorrelations of these scales. The factor loadings of the scales G, Ch, L6, N and Hy-O were .927, .868, .847, .818 and .770, respectively.<sup>34</sup>

It is claimed by Hathaway that the correction of the five clinical scales of the MMPI by use of K increases the proportion of clinically diagnosed cases scoring above the 90th percentile of normals.<sup>35</sup> Jeffery undertook an investigation of the factors which influence responses to K.

<sup>33</sup>Ibid., pp. 551-552.

<sup>34</sup>Ibid., pp. 552-553.

<sup>35</sup>S. R. Hathaway, Supplementary Manual for the MMPI (New York: Psychological Corporation, 1946).

She found that the differences characterizing high and low K were consistent for both the test and an interview situation. It was concluded that this observed set was persistent and deep-rooted. She states that the set evidences itself in items containing temporal adverbs. People who answered "true" tended to select categories containing the term "frequent" and the opposite appeared for those answering "false." Jeffery's data support the assumption that K measures two extremes of a deeply-rooted attitude causing distortion of personality items, but offered no explanation of the dynamics involved.<sup>36</sup>

McKinley, Hathaway and Meehl describe the statistical derivation of K weights as an optimal value which operates as a differential ratio in distinguishing a criterion group of abnormals diagnosed for each scale from a sample of unselected normals. The investigators emphasize, however, that this differentiation is only for the general population and state "it seems likely that for the best separation of maladjusted normals such as those which abound in a college counseling bureau and would be formally diagnosed in a psychiatric clinic as simple adult maladjustment, other weights might be better."<sup>37</sup>

A study of the K scale by Schmidt using 98 Army convalescents who were diagnosed normal found that the L scale was as good an indicator of falsifying as K. He reports that the basic shape of the profile remained the same, but its height decreases with falsification. He concluded that the K factor contributes little, if anything, to differential diagnosis.<sup>38</sup> Hunt and others tried to check the discriminatory power of the K scale using psychotic and non-psychotic adult male psychiatric patients.

---

<sup>36</sup>M. E. Jeffery, "Some Factors Influencing Answers on the Multi-phasic K Scale," (unpublished Doctor's dissertation, The University of Minnesota, 1946).

<sup>37</sup>J. C. McKinley, S. R. Hathaway and P. E. Meehl, "The MMPI: VI. The K Scale," Journal of Consulting Psychology, 1948, 12: 20-31.

<sup>38</sup>H. O. Schmidt, "Notes on the MMPI: The K Factor," Journal of Consulting Psychology, 1948, 12:337-342.

They state that the K scale did not improve diagnosis and failed to reduce "false negatives."<sup>39</sup> In another study using 109 psychology students and 74 Navy general court martial prisoners Hunt found that the K correction failed to reduce deception. However, Hunt did find that an index of F minus K correctly identified malingerers, but not faked normal profiles.<sup>40</sup> Cofer and others found K and L scores significantly higher for faked normal profiles and suggested the use of an additive combination of L and K to identify these.<sup>41</sup>

After reviewing MMPI literature Cottle concludes that K does not appear useful by itself to increase the discrimination of the clinical scales; however, the K scale in combination with L or F is useful to detect deliberate faking on the MMPI. In conclusion, Cottle states that there is evidence that the K scale reflects a persistent, deep-rooted attitude of distorting personality items and may be useful as a clinical scale to identify the defensive or the overly self-critical subject.<sup>42</sup>

Because of the contradictory evidence presented in the literature concerning the K scale and its function, no definite conclusion can be reached regarding its efficacy. Meehl and Hathaway's investigation of the K scale presented striking evidence of the validity of K to differentiate between deviant curves of actual normals and abnormals. However, in replication by other investigators the usefulness of K is not completely supported. All that can be said presently is that perhaps K or a fraction of K coupled with other validity key components can be useful in identifying certain test-taking attitudes.

---

<sup>39</sup>H. F. Hunt, et al., "A Study of the Differential Diagnostic Efficiency of the MMPI," Journal of Consulting Psychology, 1948, 12: 331-336.

<sup>40</sup>H. F. Hunt, "The Effect of Deliberate Deception on MMPI Performance," Journal of Consulting Psychology, 1948, 12:396-402.

<sup>41</sup>C. N. Cofer, J. Chance and A. J. Judson, "A Study of Malinger-ing on the MMPI," Journal of Psychology, 1949, 27:491-499.

<sup>42</sup>W. C. Cottle, "The MMPI: A Review," Kansas Studies in Education, 1953, 3:9.

### Subtle-Obvious Keys (S-O)

While the K scale was being considered and developed Wiener was developing the concept of relatively subtle and obvious keys for the MMPI scales on the same ground as those given by Meehl and Hathaway above.<sup>43</sup> It was considered that the development of such keys on individual scales of the MMPI would yield more information and be of more practical usefulness than an overall validity scale, such as L, G and finally K.

Wiener considered that extremely deviate individuals could be picked out by a test consisting of obvious items. However, to help the counselor working with a normal population, it was thought that a much more subtle test was required which would both distinguish the extreme deviates and differentiate among the characteristics of a normal population. To Wiener these two services of a personality test appeared to be best served by developing both S and O scales.<sup>44</sup>

In the development of the S and O keys the MMPI was used because 1) it was available for a relatively large normal population, 2) it was felt to have the most extensive and useful validation of any personality test, 3) it used generally accepted categories of personality characteristics, and 4) it had the unique feature of validity measures designed to indicate test-taking attitudes. It was felt, however, that when the MMPI was used with a relatively normal population, one that was functioning relatively successfully in society, certain important aspects of personality were masked because of its validation in terms of abnormal groups. The most obvious items distinguished the abnormal groups from the normal, whereas it appeared that the more subtle items should have had the greater validity in distinguishing the personality characteristics of normal groups.<sup>45</sup>

---

<sup>43</sup>D. N. Wiener, "Subtle and Obvious Keys for the MMPI, " Journal of Consulting Psychology, 1948, 12:164-170.

<sup>44</sup>Ibid., p. 164.

<sup>45</sup>Ibid.

In developing the S and O keys all items of the MMPI were divided into two groups, those to which significant responses were relatively easy to detect as indicating emotional disturbance, and those to which they were relatively difficult to detect. The items for each scale were sorted into these two categories. All F scale items that also appeared in other scales were automatically assigned to the obvious category because by definition they seldom occur in a normal population. In addition, those items for which a blank response (no check on the answer sheet) was scored in a significant direction, were assigned to the subtle keys. Pooled judgments of raters was then used to sort the other items into the two categories with no attempt made to equalize the number of items in each group. There were more obvious items than subtle. The keys thus developed were used to rescore the test sheets of a representative sampling of 100 cases of the original male norm group for the MMPI. T-scores were developed and assigned to subtle and obvious item counts on the same basis as for the total scale T-scores. Subtlety-obviousness was determined rationally and not empirically; hence S-O scales were not formed for all MMPI subtests because Wiener felt that they contained too few S items.<sup>46</sup>

Tabulation tables for the raw scores on the subtle and obvious keys indicated a positive skew for most of the O item distributions of the norm group; relatively few individuals in the normal population answered the obvious items in a significant direction. It seemed probable to Wiener that significant answers to these O items were most characteristic of an institutionalized population. The S items were distributed in a relatively normal manner.

An additional check on the validity of the selection of items for the keys was the frequency of their occurrence among the responses of a normal population. For all five scales the S items were answered in

---

<sup>46</sup>Ibid., pp. 165-166.

a significant direction approximately twice or more as frequently as the O items. The bases used to select items for the S and O keys for five scales of the MMPI yielded O items which were answered relatively infrequently in a significant direction by a normal population compared with S items, and S items whose significant answers were in a reverse direction from the expectation of both the original authors of the MMPI.<sup>47</sup>

Recent research with the subtle and obvious scales for the MMPI raises some important interpretive problems.<sup>48-51</sup> While it was not puzzling to find that hospitalized psychiatric patients and other mal-adjusted and unsuccessful groups obtained high T-scores on the obvious scales, it was disconcerting to Fricke to find that groups of recovered psychiatric patients, successful trainees, successful salesmen and college sophomores obtained higher T-scores on the S scales than the normal MMPI population. In addition, these groups obtained higher overall T-scores than groups of unrecovered psychiatric patients, unsuccessful trainees, etc.<sup>52</sup> Because each of the items in the subtle scales was originally selected by Hathaway and McKinley due to its discrimination between normal and abnormal groups, it appears that there is something common either to the items or to the groups which influences the size of the T-scores more than does the original discriminating power of the items.

---

<sup>47</sup>Ibid., pp. 166-168.

<sup>48</sup>E. Rosen, "Self Appraisal, Personal Desirability, and Perceived Social Desirability of Personality Traits," Journal of Abnormal and Social Psychology, 1956, 52:151-158.

<sup>49</sup>E. Rosen, "Self Appraisal and Perceived Desirability of MMPI Personality Traits," Journal of Counseling Psychology, 1956, 3:44-51.

<sup>50</sup>D. N. Wiener, "Selecting Salesmen with Subtle-Obvious Keys for the MMPI," American Psychologist, 1948, 3:364.

<sup>51</sup>D. N. Wiener, "A Control Factor in Social Adjustment," Journal of Abnormal and Social Psychology, 1951, 46:3-8.

<sup>52</sup>B. G. Fricke, "Subtle and Obvious Test Items and Response Set," Journal of Consulting Psychology, 1957, 21:250-252.



Fricke suggests that it is quite possible that groups obtaining high T-scores on the S scales have a response set to answer false. Fricke points out that Wiener's division of the MMPI items for five clinical scales into S and O items tends to make false the scored response for the S items. Four of the five scales have a majority of items scored false. Only the hypomania scale has more true than false responses and it is of interest to note that it is this scale which does not separate the groups studied by Wiener.<sup>53</sup>

On the basis of some data and theoretical considerations, Wiener suggests that the S items are best for assessing the personality of normal persons and that the O items are best for abnormal persons.<sup>54</sup>

Fricke states:

If it is true that S items are more likely to function in a normal population and O items in an abnormal population, then according to the present contention that K operates as a measure of response set, a difference in correlations should be reflected in normal and abnormal groups. Specifically it would be expected that the correlations of K with the five clinical scales utilized would be more positive or less negative in a normal than in an abnormal population. Correlations reveal that four of the five expectations are fulfilled; only the correlation between K and hypomania are not substantially different. A tentative hypothesis drawn from the data is that the more positive the correlation between scores from a measure of response set to answer false and scores from clinical scales composed of S-O items, the more likely it is that the group is well adjusted or successful.<sup>55</sup>

### Set T Scale

According to Cronbach who has reviewed several response sets that influence a test taker's behavior, the variance generated by a

---

<sup>53</sup>D. N. Wiener, "Subtle and Obvious Keys for the MMPI, " Journal of Consulting Psychology, 1948, 12:168-169.

<sup>54</sup>Ibid., p. 170.

<sup>55</sup>B. G. Fricke, "Subtle and Obvious Test Items and Response Set, " Journal of Consulting Psychology, 1957, 21:250-252.

response set is regarded undesirable because it contributes only error variance and cannot be used to increase the usefulness of a test.<sup>56, 57</sup> However, Fricke assumes a different rationale which stresses that response set can be used to improve the validity of many tests.<sup>58</sup>

In surveying the responses scored in most personality scales Fricke discovered that the significant responses are predominantly in one direction. For example, 96 out of the 100 items in the Cornell Index are keyed yes and 47 of 60 items in the hysteria scale of the MMPI are keyed false. A review of personality tests validated to predict academic achievement disclosed that for the more valid tests the response predictive of high achievement is usually false, no, or disagree. Some other tests in which a majority of scored responses fall into a particular response category are the California Psychological Inventory, the Humm-Wadsworth Temperament Analysis, the Minnesota Teacher Attitude Inventory, and the Strong Vocational Interest Blank.<sup>59</sup>

Fricke contends that when the number of score responses is unequally divided between all possible response categories the effect of a response set may be substantial. If response set per se is not directly related to the criterion, it operates to introduce error in the criterion predictor. Fricke states that this is usually the case, but if response set is related to another variable that is criterion-related,

---

<sup>56</sup>L. J. Cronbach, "Response Sets and Test Validity," Educational and Psychological Measurement, 1946, 6:475-494.

<sup>57</sup>L. J. Cronbach, "Further Evidence on Response Sets and Tests Designs," Educational and Psychological Measurement, 1950, 10:3-31.

<sup>58</sup>B. G. Fricke, "Subtle and Obvious Test Items and Response Set," Journal of Consulting Psychology, 1957, 21:250-252.

<sup>59</sup>B. G. Fricke, "The Development of an Empirically Validated Personality Test Employing Configural Analysis for the Prediction of Academic Achievement," (unpublished Doctor's dissertation, University of Minnesota, 1954).

then response set may be used as a suppressor variable.<sup>60</sup> By suppressing or removing the influence of response set from the criterion-predictor an improvement in test validity can be effected.<sup>61-64</sup>

Fricke originally held the view that the imbalance of true and false items was simply a function of how the statements were written. But this seemed unlikely since on some tests having scales for several personality dimensions, the scored direction is different for different scales. It was concluded that the nature of the criterion groups and not the items was responsible for the imbalance. As a result Fricke set out to determine whether or not a measure of response set could be used to increase test validity.

One index or measure of set for answering true was obtained by counting the times a person marked true to statements in a test. This method was utilized by Humm and Wadsworth to get a measure of suggestibility or cooperativeness. They counted the times a test taker answered no to questions in the Temperament Schedule. Answer sheets having a large or small "no count" were not considered sufficiently valid for further analysis. According to their suggested cut-off points, 30-60 percent of the answer sheets were rejected as invalid.<sup>65</sup>

Fricke considered that a much more sensitive index of response set would be a count of the true responses for those statements which

---

<sup>60</sup>Ibid.

<sup>61</sup>Paul Horst, "The Prediction of Personal Adjustment," Social Science Research Council Bulletin, 1941, Number 48.

<sup>62</sup>A. S. Levine, "A Technique for Developing Suppression Tests," Educational and Psychological Measurement, 1952, 12:313-315.

<sup>63</sup>Quinn McNemar, "The Mode of Operation of Suppressant Variables," American Journal of Psychology, 1945, 58:554-555.

<sup>64</sup>R. J. Wherry, "Test Selection and Suppressor Variables," Psychometrika, 1946, 11:239-247.

<sup>65</sup>D. G. Humm and G. W. Wadsworth, "The Humm-Wadsworth Temperament Schedule," American Journal of Psychiatry, 1935, 92:163-200.

49-51 percent of the test-takers marked true. For such statements there was no general agreement on an answer; they held maximum controversiality. A person without a response set would respond true as often as false on these items. A person with a high score could be thought to have a strong set for marking true. A low score would indicate the opposite. Fricke emphasized that items for a set scale would be selected independently of any criterion external to the test. Because it was difficult to obtain enough items to form a scale at the 49-51 percent level of controversiality, it was decided to accept for the true response set scale (Set T) those statements in the Opinion, Attitude and Interest Survey which 40-60 percent of each criterion group marked true.<sup>66</sup> Hence, the Set T scale of the OAIS consisted of 69 statements scored in the true direction.

Upon completion of his investigation, Fricke concluded that a measure of response set could be used to increase greatly the accuracy of criterion prediction.<sup>67</sup>

#### The B Scale

The B scale of the MMPI is structurally and functionally similar to the K scale but differs markedly from K in the method used for its construction. B, a measure of response bias, was modeled after the Set T scale of the OAIS.<sup>68</sup> The B scale and the Set T scale made it possible to measure a test-taker's tendency to answer true to statements in a personality inventory. Individuals with a strong bias to answer true

---

<sup>66</sup>B. G. Fricke, The Opinion, Attitude and Interest Survey (Minneapolis: Investors Diversified Services, 1955).

<sup>67</sup>B. G. Fricke, "Response Set as A Suppressor Variable in the OAIS and MMPI," Journal of Consulting Psychology, 1956, 20:161-169.

<sup>68</sup>B. G. Fricke, The Opinion, Attitude and Interest Survey (Minneapolis: Investors Diversified Services, 1955).

obtain low hysteria scores on the MMPI due to the fact that 78 percent of the Hy items are scored false.<sup>69</sup> Fricke felt that if a fraction of a Set T type of scale was added to Hy or if a fraction of K was subtracted from Hy, the validity of Hy would be improved. The assumption was that by partialing out, or suppressing the influence of response bias, the purity of the clinical scale would be increased.<sup>70</sup>

To select items which would reflect response bias on the MMPI Fricke examined the item responses of normal persons. His objective was to obtain a pool of items of high controversiality, that is, items to which about equal numbers answer true and false. Items drawing true answers from 40-60 percent were considered sufficiently sensitive to be useful. Since some normals did not answer true or false to every item, half the "cannot say" items were added to the true answers to establish whether or not each item met the arbitrary 40-60 percent level of controversiality. Fricke considered that a test-taker with a strong bias to answer true would be expected to achieve a high score when all the controversial items were scored.

Two normal samples were involved. The first sample of 604 cases was used by Hathaway and McKinley in the construction of the clinical scales; it consisted of a sub-group of 339 Minnesota normals and a sub-group of 265 college normals. The percentage of these two sub-groups who answered true to each item was averaged. The second sample of 589 cases was used as the norm group for the more recently constructed non-clinical scales; it consisted of a sub-group of 253 normal males and a sub-group of 336 normal females. The percentages for these two sub-groups also were averaged. A total of 81 items were located which were answered true by 40-60 percent of both normal

---

<sup>69</sup>B. G. Fricke, "Conversion Hysterics and the MMPI, " Journal of Clinical Psychology, 1956, 12:322-326.

<sup>70</sup>B. G. Fricke, "A Response Bias Scale for the MMPI, " Journal of Counseling Psychology, 1957, 4:149-153.

samples. Of the 81 items of high controversiality 18 were found to be in the 30-item K scale. The response bias scale B consists of the 63 non-K items. Fricke states that high T-scores on B indicate a tendency to answer false.

Because of the complete lack of evidence in the literature concerning Fricke's interpretation of the B scale, no definite conclusion can be reached regarding the efficacy of the B scale.

### Miscellaneous Scales

The G scale. About three years before research on the test-taking attitude was begun, Hathaway and Estes, using a variant of the method of internal consistency, developed a scale called G. This scale was the only MMPI scale which was derived without the use of a criterion external to the test; the selection and scoring of item was based wholly upon the intercorrelations among the items themselves. Essentially, the procedure consisted in locating among a group of 101 unselected normals those individuals who, when their answer sheets were used as scoring keys, produced the maximum variance of the other 100 scores. The assumption was that these persons were the most extreme deviates on whatever factor or factors contributed most heavily to the variance and covariance of the total pool of MMPI items. From the evidence adduced by Mosier, it is of course clear that the purity of factorial unity of this hypothetical underlying continuum is by no means guaranteed by such a procedure.<sup>71</sup> Meehl and Hathaway state that this maximizes the variance of a set of items by scoring them in such a direction as to maximize their mean covariance (since the item variances are unaffected by the direction of scoring).<sup>72</sup> Instead of actually calculating the variances for the  $2^{50}$  ways

---

<sup>71</sup>C. I. Mosier, "A Note on Item Analysis and the Criterion of Internal Consistency," Psychometrika, 1936, 1:275-282.

<sup>72</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:549-550.

of scoring the test, the investigators selected individuals who approximated the optimal scoring key. It was found that the scoring keys for some ten individuals selected by this method tended to form two distinct clusters, each of which consisted of keys (individuals) showing high correlations with one another and high negative correlations with the members of the other cluster. An item analysis was then carried out on these two small groups, and the items resulting were combined into a scale called G (general factor).

The G scale, although derived without recourse to any clinical group whatever, nevertheless showed a correlation of .91 with clinical scale Pt. The mean MMPI curves for unselected normals with high G (the neurotic end) showed elevations on seven MMPI clinical scales and on F; whereas L (raw score) and Hy tended to fall below the mean. The mean profile for normals with low G was almost an exact mirror image of this curve. However, G was not found to be effective in the detection of any clinical group or to be particularly useful for any purpose; and since at that time no theoretical basis was available for interpreting it, the scale was abandoned. Another scale, called plus was derived in a similar but not identical manner.<sup>73</sup>

The N scale. When Meehl and Hathaway first discovered that certain clearly abnormal persons obtained normal MMPI profiles, and on the other hand, that certain normals obtained elevated profiles with no evidence of deliberate falsification, they began an investigation which initially led to the development of the N scale. The key discriminated individuals who were overly critical in reporting themselves (plus-getters) from actually abnormal subjects with similar MMPI profiles. This scale, however, did not prove to be useful in detecting impulsive or defensive sortings. For this reason, further study was carried out which resulted ultimately in the development of K.<sup>74</sup>

---

<sup>73</sup>Ibid., p. 550.

<sup>74</sup>S. R. Hathaway and J. C. McKinley, "A Multiphasic Personality Schedule: I. Construction of the Schedule," Journal of Psychology, 1940, 10:249-254.

The Ch scale. In the derivation of the original hypochondriasis key, there was developed a correction scale called Ch, the function of which was to separate actual clinical hypochondriacs from a group of non-hypochondriacal abnormals who attained spuriously elevated scores on H. The item content of this Ch key was quite puzzling to Hathaway and McKinley because, although the correction was successful, the items did not appear to refer to anything either hypochondriacal or anti-hypochondriacal. They possessed no apparent psychological homogeneity. The majority of the items on Ch were scored if answered in the statistically rare and obviously maladjusted direction. They apparently measured some non-somatic component of test responses which resulted in spuriously elevated H scores in persons who were not actually hypochondriacal.<sup>75</sup>

The Hy-O scale. The Hy-O items reflected the N component, although they were scored in the opposite direction from N, as well as from Ch and G. Hy-O consist of a sub-set of items which were chosen because they differentiated empirically the hysteria clinical group from normals. It was concluded that these items reflected the self-deceptive and impunitive attitude of the hysterical temperament.

The Cd and the Hy-S keys. Other minor attempts of validity key construction for the MMPI consists of the Cd and Hy-S scales. The former was developed to differentiate systematic depressives from normals; the latter items constituted an attempt to distinguish hysteria, hypomania and psychopathic deviates from normals.<sup>76, 77</sup>

<sup>75</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI, " Journal of Applied Psychology, 1946, 30:550-551.

<sup>76</sup>S. R. Hathaway and J. C. McKinley, "A Multiphasic Personality Schedule: III. The Measurement of Symptomatic Depression, " Journal of Psychology, 1942, 14:73-84.

<sup>77</sup>J. C. McKinley and S. R. Hathaway, "A Multiphasic Personality Schedule: V. Hysteria, Hypomania and Psychopathic Deviate, " Journal of Applied Psychology, 1944, 28:153-174.



Consistency keys. Consistency or verification of response checks are no longer infrequent keys in various objective personality and interest inventories. The instruments devised by Kuder, Strong and Edwards typify contemporary attempts to uncover falsification by using consistency of response to determine whether the subject understood test directions, possessed sufficient reading skills to comprehend the test items, or answered carelessly or insincerely.<sup>78-80</sup>

### Similarity Between Various Scales

B and K scales. Although the methods used in construction of the B and K scales were vastly different, structurally and functionally B and K are quite similar. Both consist of items of high controversiality and both have their items scored in one direction.<sup>81</sup>

The functional similarity of B and K is revealed by the correlation coefficients shown in Table IV. The correlation between B and K for 336 normal females is minus .68 and the correlation for 253 normal males is minus .67. The correlation for a sample of 63 conversion hysterics was found to be minus .73.<sup>82</sup>

Fricke concluded from the data in Table IV that B and K were the two most similar validity indicators. It is of interest to note that the correlations between K and L are higher than the correlations

---

<sup>78</sup>G. F. Kuder, Kuder Preference Record: Vocational (Chicago: Science Research Associates, 1956), p. 3.

<sup>79</sup>G. F. Kuder, Kuder Preference Record: Occupational (Chicago: Science Research Associates, 1959), p. 4.

<sup>80</sup>A. L. Edwards, Edwards Personal Preference Schedule (New York: The Psychological Corporation, 1959), p. 15.

<sup>81</sup>B. G. Fricke, "A Response Bias Scale for the MMPI," Journal of Counseling Psychology, 1957, 4:150.

<sup>82</sup>Ibid., p. 151.

between B and L and that the correlations between B and F are higher than the correlations between K and F. From this Fricke interpreted that the L component was stronger in K than in B, and the F component was stronger in B than in K.

Table IV. Intercorrelations of Four MMPI Validity Scale Indicators and Hy for Normals

	Female		Male		
	L	F	K	B	Hy
L		-.28	.38	-.30	.13
F	-.01		-.36	.40	.19
K	.36	-.35		-.68	.29
B	-.28	.43	-.67		-.23
Hy	.19	.04	.39	-.37	

The correlations between Hy and B, and between Hy and K suggest that B and K could be used as suppressor variables. Fricke suggests that the Hy scale could be improved by adding to it a certain fraction of B or by subtracting from it a certain fraction of K. However, what values should be used were not discussed.<sup>83</sup>

B and K as discriminators. A function more influential than the suppressor action is evidently operating in B and K. Fricke states that B and K are both unsuccessful suppressors because they both discriminate conversion hysterics from normals; a low B and a high K are indicative of conversion hysteria. Consequently, this investigator concludes that the subtraction of a fraction of B from Hy and the addition of a fraction

---

<sup>83</sup>B. G. Fricke, "Subtle and Obvious Test Items and Response Set," Journal of Consulting Psychology, 1957, 21:250-252.

of K to Hy would improve the validity of Hy if B and K tap something diagnostic of conversion hysteria that is not tapped by Hy.<sup>84</sup>

It is important to note that B and K each appear to function simultaneously as scale suppressor and criterion discriminator. This is unfortunate since the suppressor and discriminator effects are in opposite directions and tend to cancel each other. The influence of B and K as suppressors is weaker than their influence as discriminators and this results finally in the subtraction of B and the addition of K.<sup>85</sup>

Because the level of K scores, and probably B, is affected by the social-educational-economic level of the test-takers, it was Fricke's speculation that for the lower levels the discriminator role of K (and B) would be much more important than the suppressor role, but that for the higher levels (college students) the discriminator role of K and B would be much less important than the suppressor role. A possible explanation for McKinley, Hathaway, and Meehl's finding that the addition of a fraction of K to Hy did not improve Hy is to be found here. If their conversion hysterics and normal control cases had the same mean K scores, then the addition of a fraction of K to Hy would not improve its validity; subtraction of a fraction of K probably would have capitalized on K's suppressor role and improved the validity of Hy.<sup>86</sup>

Discrepancies in B and K scores may be of some importance (that is, it is possible that a test-taker with a T score of 70 on B and a T score of 60 on K is less defensive than a test-taker with a T score of 50 on B and a T score of 60 on K; it might be argued that the first test-taker's K score was obtained largely due to his response set to answer false but that the second test-taker's K score was obtained through the operation of something other than response set).

---

<sup>84</sup>B. G. Fricke, "A Response Bias Scale for the MMPI, " Journal of Counseling Psychology, 1957, 4:152.

<sup>85</sup>Ibid.

<sup>86</sup>Ibid.

The assumption is that K is more than a measure of response set. If the assumption is not valid, there probably is no need for the two test scores.<sup>87</sup>

While both scales appear to reflect to a certain extent a test-taker's nontest behavior, their primary role is to reveal something about a test-taker's test taking behavior so that more accurate inferences can be drawn from the diagnostic clinical scales.

Functional similarity of K and Set T. The functional similarity of K and Set T scales is seen in the fact that two-thirds of all K items fall in the controversial range. Further evidence on the similarity of Set T and K is found in their correlation with each other: minus .58 to minus .71. To Fricke, it appeared that Set T and K accomplished essentially the same thing. Whether Set T was superior to K was not determinable from the available data. While K may not be as pure a measure of response set, it may do something in addition to what is done by Set T.<sup>88</sup> Fricke concludes:

A measure of response set such as K or Set T may function as a suppressor when there is no true-false imbalance, and may not function as a suppressor when there is a moderate true-false imbalance. Too frequently scores from the K scale have been correlated with scores from other tests, and judgments have been made as to what K was "really" measuring. I don't know what K measures, no one else does either. It seems unlikely that response set keys would be able to bring validity to tests which are not validated empirically even though response set may be a major component in a test taker's score.<sup>89</sup>

The marked structural and functional similarity of Set T of the OAS and K of the MMPI was drawn upon to challenge the traditional interpretation of the K scale. Evidence was assembled which indicates that some of the MMPI scales are not optimally K corrected.

---

<sup>87</sup>Ibid.

<sup>88</sup>Ibid.

<sup>89</sup>Ibid.

S and O relationships. Relationships exist between the S and O, and the L scale, intelligence, psychological sophistication and neuropsychiatric diagnosis. The group with high scores on the L scale of the MMPI was higher on the S keys of all five scales than on the O keys, and was also higher on the S keys than was the low L scale group. For the group with low L scores, the O scores were for all scales approximately equal to or higher than the S scores. Individuals of high ability (intelligence T score above 60 on the Otis) have approximately equal O and S scores, whereas individuals of low ability (T score below 40) have generally higher O scores than S, and higher O scores than the high ability group.<sup>90</sup>

MMPI profiles of a psychologically sophisticated group showed distinction between S and O keys, with S much higher than O whether the group was giving honest results or was attempting to fake good. With this sophisticated group it appeared to make little difference whether the test was taken honestly or faked good; in either case, O items were successfully avoided whereas S items yielded average and above average T scores.<sup>91</sup>

In general, O keys are highly correlated with each other and have no correlation with S keys; S keys have a low positive correlation with each other. There is a high negative correlation between O minus S and K, indicating the considerable weighting of K with S items. There is evidence that high L scale scores are associated with higher S than O scores, whereas the converse is true for low L scores; psychologically sophisticated individuals almost completely avoid significant O responses and have much higher S scores. While total scale scores on the MMPI failed to differentiate significantly between successful and unsuccessful students and on the job trainees, O keys were

---

<sup>90</sup>D. N. Wiener, "Subtle and Obvious Keys for the MMPI," Journal of Consulting Psychology, 1948, 12:169.

<sup>91</sup>Ibid., p. 170.

significantly higher than the S for the unsuccessful group. The S keys were insignificantly lower, and the total scale scores were between the two.<sup>92</sup>

### Summary

All major validity scales have been examined in this chapter. The order of discussion of the various keys was based upon the chronology of their development with the earliest devised scales discussed first. In addition to the major scales, however, several minor scales had to be considered in order to bring about a more lucid understanding of the major keys. All of the scales were developed around the MMPI with the exception of Set T. Set T was constructed with OASIS data.

Essentially, 1) the L scale used items which were socially desirable but rarely true of an individual; 2) the F scale was developed with rarity items (items chosen by ten percent or fewer of the sample); 3) the K scale items were selected because they differentiated normal individuals who scored abnormally from abnormal subjects who scored normally; 4) the S-O keys distinguished abnormal subjects from normal by the use of a subtlety-obviousness continuum; and 5) Set T and B scale items were developed by uncovering high controversial items and measured response set.

Clear probability statements concerning the validity of the various scales are lacking.

The motivational research study with its instrumentation and the procedures utilized in the present investigation are discussed in Chapter III.

---

<sup>92</sup>Ibid.

## CHAPTER III

### PROCEDURES

The present investigation used the administered test protocols of participants in Farquhar's motivational research project.<sup>1</sup> In this chapter the background, theory, design and instrumentation of Farquhar's project are discussed and the procedures for the present study are outlined.

#### Background of Farquhar's Study

In an attempt to collate a thorough objectively validated description of the personal characteristics of high and low academically motivated students, Farquhar scrutinized the existing literature on motivation. As a result, a theory of need-achievement and non-need-achievement motivation was formulated. A summary of the basic motivational theory is found in Tables V and VI.

Table V. Summary of Theory of Need-Achievement and Non-Need-Achievement Motivation Basic to Current Research

Motivational Situation	
Need-Achievement	Non-Need-Achievement
1. Long term involvement	1. Short term involvement
2. Competition with a maximal standard of excellence	2. Competition with a minimal standard
3. Unique accomplishment	3. Common accomplishment

<sup>1</sup>William W. Farquhar, "A Comprehensive Study of the Motivational Factors Underlying Achievement of Eleventh Grade High School Students," (East Lansing: Awarded Research Application to the Commissioner of Education, United States Office of Education, November 1, 1959), 15 pp. (Mimeographed.)

Table VI. Hypothesized Personality Factors Associated with Academic Achievement

Academic Anxiety	Independence-Dependence Conflict
Self Value	Activity Patterns
Authority Relations	Goal Orientation
Interpersonal Relationships	

Three attitudinal areas were ascertained which were considered capable of differentiating between over-and under-achievers. These areas consisted of attitudes toward school and learning, toward self and toward parents. Later, another area was established: attitudes toward occupations.

In order to test the posited motivational theory 725 items were logically developed with the purpose of contributing significantly to the regression of an academic predictor on grade achievement. The essential assumption was that students who exemplified extremes in academic performance should also exhibit extreme response to motivational characteristics.

#### General Design of the Motivational Study

The above hypothesis was tested with a sample of approximately 4200 eleventh grade Michigan public high school students who were attending nine separate institutions. Under-and over-achieving students were identified by the following procedure: 1) Schools which had 9th grade Differential Aptitude Test scores available on their current 10th graders were contacted and asked to cooperate in the study. 2) A second aptitude measure was obtained so that reliable estimates of academic aptitude could be made. California Tests of Mental Maturity were administered to schools lacking this data. 3) The DAT-Verbal Reasoning subtest and the CTMM Language scores were used in obtaining a stable estimate of



academic aptitude after empirically examining possible DAT CTMM sub-score combinations. 4) Regression lines were calculated for each school and sex assuming a perfect correlation between DAT-VR and CTMM-L sub-tests. Separate equations were calculated because a pilot study indicated that one equation could not be generalized from school to school. Only those individuals who fell within one standard error of estimate above and below the regression line were included in the study. The methodology of selection of individuals with stable measured aptitude is summarized in Figure I. Because it was important that the criterion groups be classified with little chance of making a Type I error (accepting instead of rejecting) it was decided to run the risk of Type II error (rejecting instead of accepting) even if sample were lost in the process.

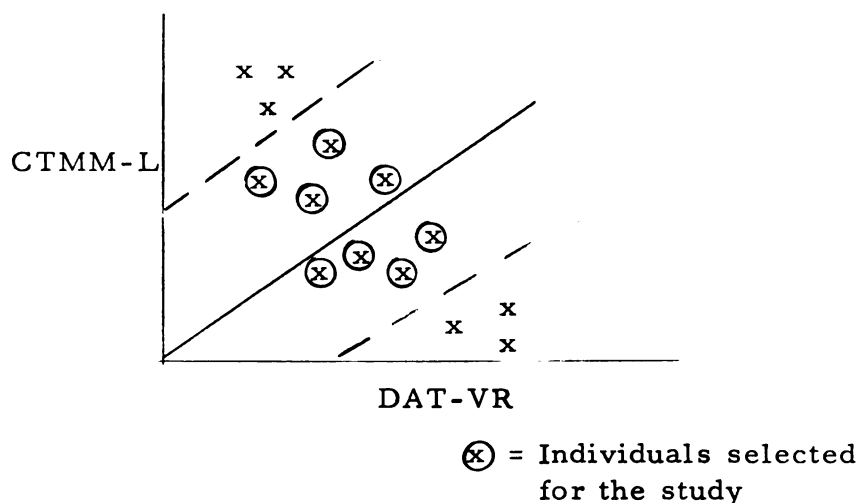


Figure I. Methodological Selection of Individuals with Stable Measured Aptitude.

5) Regression equations predicting grade point average from DAT-VR scores were calculated for each sex in each participating school. The DAT-VR was used because it was found to correlate consistently higher with grade-point average than the CTMM-L scores. Under-achievers were defined as those individuals whose grade point average fell at least one standard error of estimate below the regression line prediction of

achievement. Similarly, over-achievers were designated as falling one standard error above the regression line. The methodology used in selecting under-and over-achievers is summarized in Figure II.

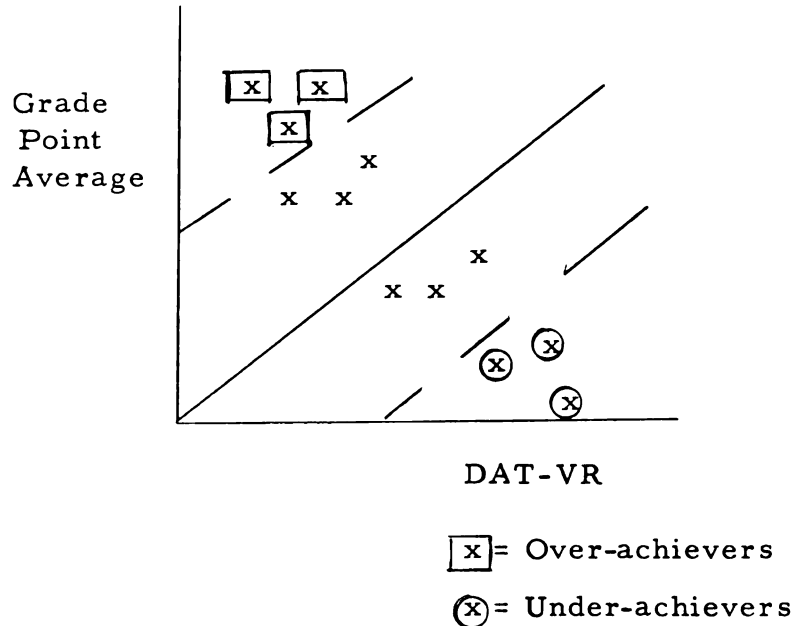


Figure II. Method of Selecting Under-and Over-Achievers.

By using the above method under-and over-achievers were selected from the full range of academic aptitude. Approximately twelve percent of the sample was classified in one of the extreme groups.

### Instrumentation

The Generalized Situational Choice Inventory was developed by reviewing studies of over-and under-achieving students to determine characteristic differences. These characteristics were incorporated in pairs of statements which typified need or non-need-achievement. Classification into need or non-need-achievement was based on face validity of the content of choice. Choices had to meet the following criteria: 1) be as free as possible of culture stereotypes; 2) allow the

individual to project his choice into the future; and 3) be purely as possible need-achievement or non-need-achievement. It was not necessary that the choice be concerned with only one of the three sub-factors provided it could be clearly designated need or non-need achievement. The inventory consists of 200 item pairs.

The Preferred Job Characteristics Scale was developed essentially in the same manner as the inventory above. Previous studies indicate that occupational goals are positively correlated with need-achievement.<sup>2</sup> The scale consists of 64 items.

The Human Trait Inventory was developed by reviewing items from personality measures which previous research found to differentiate between under- and over-achieving students. The inventory consists of 120 items.

The Word Rating List was developed by using adjectives which described under- and over-achievers. The inventory consists of 110 items.

### Procedures for the Present Investigation

#### Selection of the F Scale

The F scale was selected for development because it appears to bring into consideration more invalidating variables which influence test interpretation than any other single validation scale. It is postulated that an F scale can differentiate protocols of test-takers 1) who are uncooperative; 2) who can not comprehend the test items; 3) who make clerical errors; and 4) who intentionally place themselves in a bad light. In addition, Meehl and Hathaway present evidence that the F scale can differentiate between faked and non-faked protocols.<sup>3</sup> (The rationale for selection of F is presented on page 66.)

---

<sup>2</sup>Ibid.

<sup>3</sup>P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI, " Journal of Applied Psychology, 1946, 30:536-537.

### F Scale Development

F scale items comprise those statements chosen with relatively low frequency by each sex for each of the four above inventories. The criterion for rarity of response for this study was placed at ten percent or less of the sample answering in one direction.

F items were determined first on a stratified random sample consisting of 132 males and 132 females. Stratification was based upon the identified over-achievers, under-achievers, and normals in the nine Michigan public high schools in Farquhar's motivational study. Under- and over-achievers were chosen in direct proportion to their representation in the pilot distributions. Both the over-achieving and under-achieving samples for both sexes consisted of sixteen subjects each. The subjects making up each of these groups represent approximately twelve percent of the total sample. This percentage was found by Farquhar to represent the frequency of membership in any of the extreme motivational groups in the 4200 sampling of the original study. The normal sample consisted of 100 males and 100 females.

An identical procedure was followed with a cross validation sample of equal numbers. Only those F items which were found to be in common between validation and cross validation groups were included in the final F scale. The completed F scale is based upon a sample of 264 males and 264 females.

### Sex Differences in F Item Selection

Sex differences in F item selection (that is, avoidance in selection of the item) were graphically analyzed. F items which were selected by males only, by females only and by both sexes jointly were determined.

### Reliability of the F Scale

Hoyt's analysis of variance technique was used in determining

F scale reliability.<sup>4</sup> Stratified random samples of 66 males and 66 females, treated separately, were used in determining F item reliability.

#### Validation of the F Scale

It was assumed that under-achievers could be differentiated from over-achievers by F item selection. Because the F scale is a measure of social conformity (due to the 90% criterion for selection of the item), the over-achiever was expected to select few F items. Conversely, the under-achiever was expected to select significantly more F items because of the non-conformity characteristics of his behavior.

The following procedure was used to validate the Generalized Situational Choice Inventory F scale:

- 1) GSCI items which differentiated between over- and under-achievers were determined for the 171 male over-achievers and for the 137 male under-achievers.
- 2) An item frequency distribution was constructed for each group. The point of overlap where under-achievers scored as over-achievers on GSCI items was determined by plotting the respective normal distribution curves. The point of overlap which was to be identified is illustrated in Figure III.

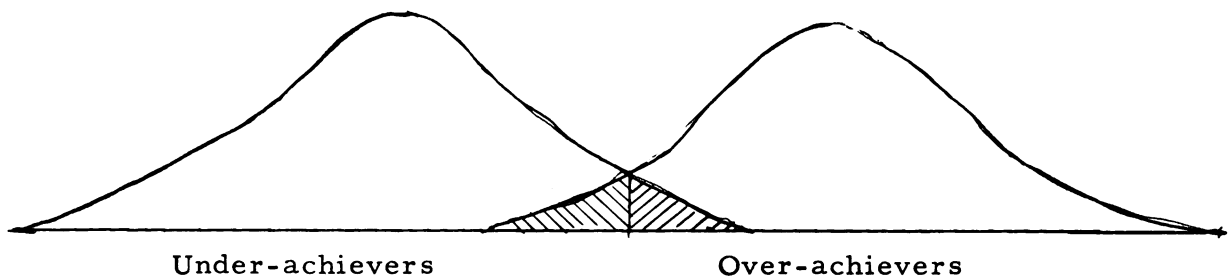


Figure III. Theorized Model of Selection of Misclassified Under-Achievers Differentiated by the GSCI.

---

<sup>4</sup>C. J. Hoyt, "Test Reliability Estimated by Analysis of Variance," Psychometrika, 1941, 6:153-160.

- 3) Under-achievers who possessed GSCI scores as large as or greater than the overlap point were identified. These under-achievers scored as over-achievers on GSCI discriminating items. In addition, the over-achievers who selected less than the overlap point scored as under-achievers on GSCI discriminating items. The above groups were then labeled misclassified over-achievers and misclassified under-achievers, respectively.
- 4) Two random samples of equal numbers were selected from the remaining identified over-achievers and under-achievers. These groups were labeled properly classified over- and under-achievers.
- 5) An F item distribution for all four inventories was constructed for the misclassified and properly classified samples. The point of overlap where misclassified under-achievers scored as properly classified under-achievers was determined by plotting the respective normal distribution curves for the F items. An identical procedure was followed using misclassified over-achievers and properly classified over-achievers.
- 6) Replication of the above procedure was carried out by randomly assigning the misclassified and properly classified individuals into original and replicative samples. Two properly classified and two misclassified F item distributions were obtained. The common point of overlap between the original and replicative groups determined the critical F score. The selection procedure for determining the F item overlap score is illustrated in Figure IV.

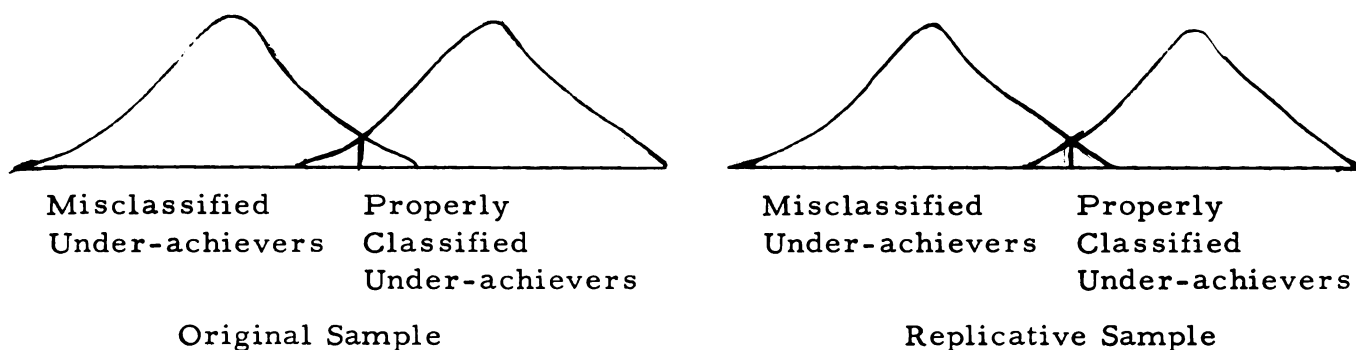


Figure IV. Selection Procedure for Determining F Item Overlap Score.

- 7) To obtain evidence of F scale validity three approaches were examined for the effect of F on: 1) expectancy of response fake; 2) test reliability; and 3) test validity.
- 8) t-tests were then used to determine significant differences between F item means of the over-achieving misclassified and properly classified groups as well as of the under-achieving misclassified and properly classified samples.
- 9) It was hypothesized that the GSCI reliability would increase after the application of the F scale because homogeneity of test performance would be increased. The effectiveness of the F scale would be determined by its ability to remove unstable individuals.
- 10) The critical F score was then applied to a stratified random sample of 66 over-achievers, under-achievers and normals. Individuals possessing total F scores as large as or greater than the critical F score were then excluded from the sample.
- 11) Reliabilities of the GSCI discriminating items were then estimated before and after application of the F scale through Hoyt's analysis of variance technique.
- 12) It was hypothesized that the validity correlation of GSCI raw scores and grade point average would increase after the application of the F scale. Again, the effectiveness of the F scale would be determined by its ability to remove unstable individuals.
- 13) An identical procedure was followed using the 189 female over-achievers and the 173 female under-achievers.
- 14) An F item distribution was constructed using the cross validation sample of 132 males and 132 females.

#### Rational of High Fake Expectancy

Through the findings disclosed in Farquhar's study over-achievers tend to be highly stable, social conforming individuals. Conversely, under-achievers tend to be more erratic, less conforming persons. In addition, under-achievers tend to avoid "risk" situations which might possess the potential of placing them in a bad light. It is likely then

that the under-achiever does not want to appear in a bad light test-wise and is more erratic in test behavior. Therefore, the under-achiever would tend to select more non-conforming or F scale items than the over-achiever.

### Summary

In this chapter the background, theory, design and instrumentation of Farquhar's motivational project were discussed. In addition, the design of the present investigation was outlined. The results of the study are discussed in Chapter IV.



## CHAPTER IV

### RESULTS OF THE INVESTIGATION

The outcomes of the present investigation are discussed in this chapter. The selection of items comprising the F scale, sex differences in F item selection, reliability and validation of the scale are presented.

#### Selection of the F Scale

Rarity responses (based upon a 10% or less criterion for selection of the item) were determined separately for a stratified random sampling of 132 males and 132 females. Thirty-five rarity items were selected by the males from the total test battery. Eighty-six rarity items were chosen by the females from the total battery.

The cross validation stratified random sample of 132 males and 132 females selected from the total test battery forty-one and ninety-two rarity items, respectively.

Only those rarity items which were commonly selected by both the validation and the cross-validation samples comprise the F scale. In the final form the male F scale consists of twenty-five F items. The female F scale is composed of seventy-three F items. The number of rarity items appearing in each of the four tests for both sexes for validation and cross validation samples appears in Table VII.

Approximately 90% of the items which were not supported under the cross validation process would have been in common agreement if the item selection criterion had been placed at 15%.

Table VII. Number of Rarity Items in Each Test

Test	Validation Sample		Cross Validation Sample		Final F Scale	
	Males -	Females	Males -	Females	Males -	Females
GSCI	10	27	11	34	6	24
Human Trait	10	11	12	17	8	11
Word Rating List	10	27	14	29	8	26
Preferred Job Characteristics	5	21	4	12	3	12
Total	35	86	41	92	25	73

Sex Differences in F Item Selection

Sex differences in F item selection are illustrated in Table VIII.

Table VIII. Sex Differences in F Item Selection

<u>GSCI</u>			<u>Word Rating List</u>		
Males	6		Males	8	
Females	24		Females	26	
Overlap	4		Overlap	8	
<u>Human Trait</u>			<u>Preferred Job</u>		
Males	8		Males	3	
Females	11		Females	12	
Overlap	5		Overlap	3	

Males selected five rarity items which were not chosen by females. Females selected fifty-three rarity items which were not chosen by males. Both males and females selected in common twenty F items.

Distribution of F Items

An F item frequency distribution was constructed for the cross validation sample of 132 males and 132 females. Of a total of 25 possible F choices 37% of the males selected no F item. Eighty-seven percent of the sample chose four or less rarity responses and ninety-four percent selected six or less items. The highest number of F items selected by any one individual was sixteen. The male F item frequency distribution is illustrated in Figure V.

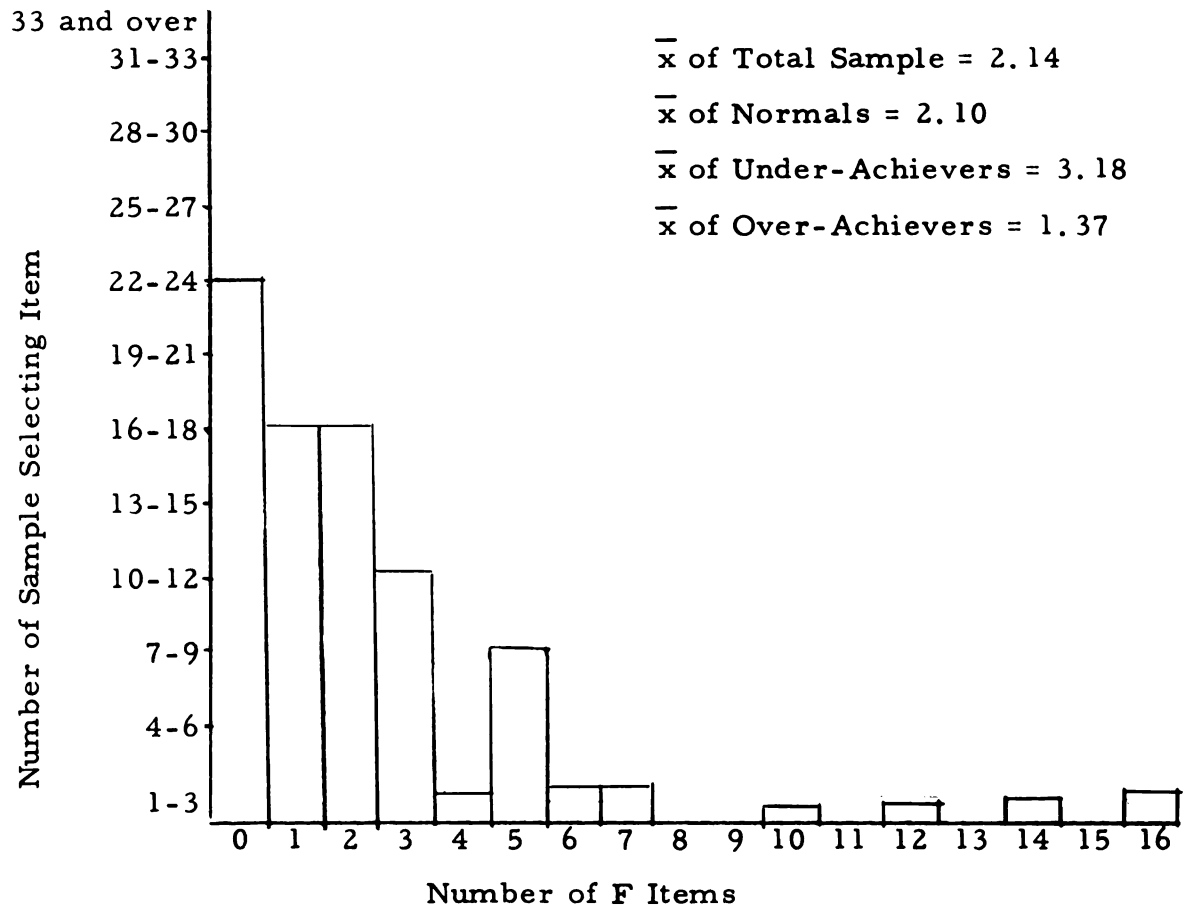


Figure V. Cross Validation F Item Frequency Distribution for Males (N = 132)

Of a total of 73 possible F choices for the females eleven percent of the sample selected no F item. Eighty-four percent of the sample chose eight or less rarity responses and ninety-one percent selected ten or less items. The highest number of F items selected by any one individual was 28. The female F item frequency distribution is illustrated in Figure VI.

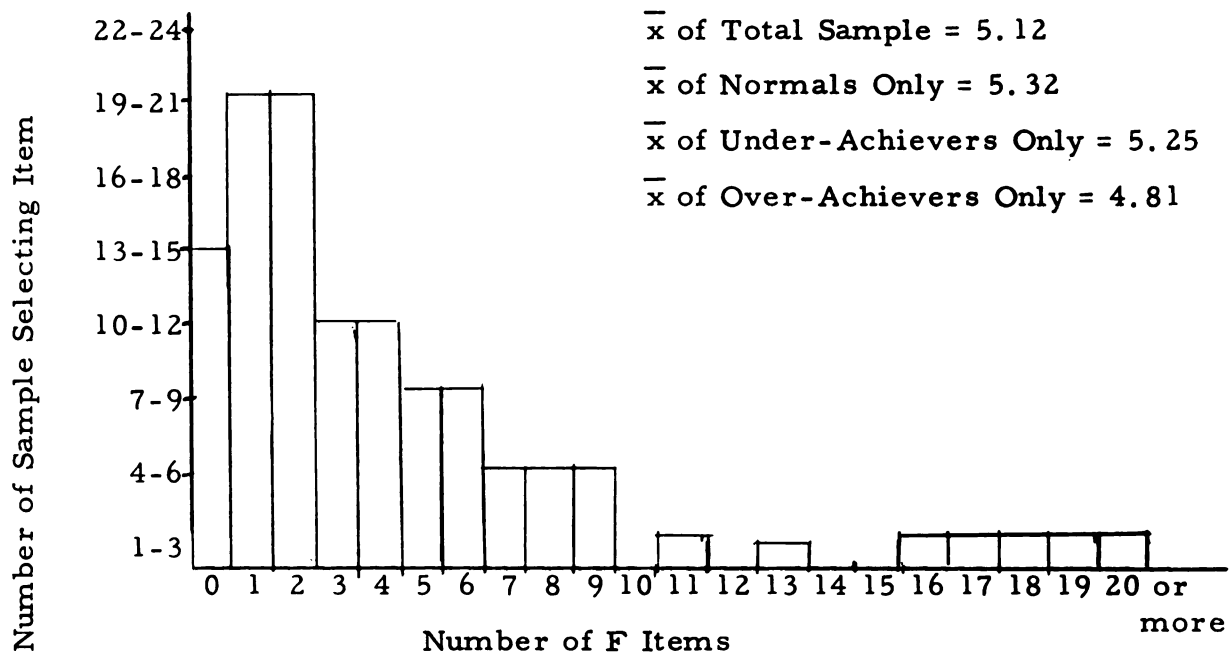


Figure VI. Cross Validation F Item Frequency Distribution for Females (N = 132)

#### Reliability of the F Scale

Hoyt's analysis of variance technique was used in estimating F scale reliability. Reliability was based on a stratified random sample of 66 males and 66 females. For the males a reliability estimate of .729 was obtained. For the females a reliability estimate of .746 was determined.

### Validation of the F Scale

To obtain evidence of F scale validity three approaches were examined for the effect of F on: 1) expectancy of response fake; 2) test reliability; and 3) test validity.

#### Effect of F on Expectancy of Response Fake

It was assumed that under-achievers could be differentiated from over-achievers by F item selection. Because the F scale is a measure of social conformity (due to the 90% criterion for selection of the item), the over-achiever was expected to select few F items. Conversely, the under-achiever was expected to select significantly more F items because of the non-conformity characteristics of his behavior.

#### GSCI Overlap

The GSCI was used to obtain evidence of the validity of the F scale. This instrument was used because it embodied to a greater extent than the other battery instruments the motivational theory of Farquhar's project.

Forty-five GSCI items differentiated at the 10% or better level of confidence after cross-validation between the 171 male over-achievers and the 137 male under-achievers. For the female sample (consisting of 189 over-achievers and 173 under-achievers) thirty items were established. An item frequency distribution was constructed for each group for each sex. The point of overlap where under-achievers scored as over-achievers on GSCI discriminating items was determined by plotting the distribution curves. The overlap point was identified as thirty-one GSCI items for the males and nineteen items for the females. The overlap points are illustrated for each sex in Figure VII.

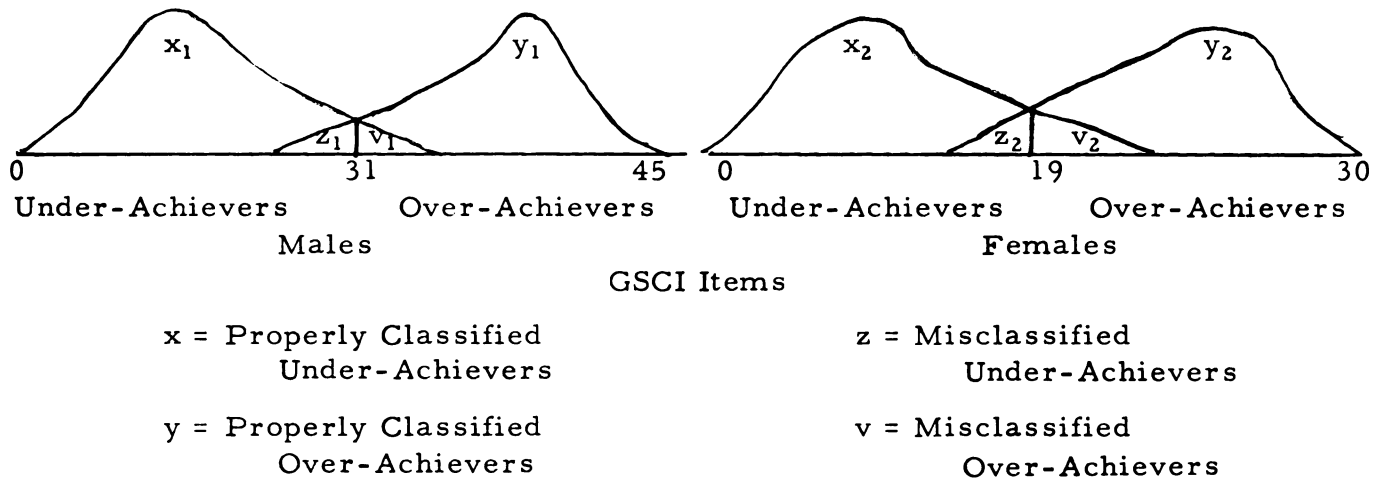


Figure VII. GSCI Overlap Points for Over- and Under-Achievers for Each Sex

Under-achievers who selected as many as or greater than the overlap point scored like over-achievers on GSCI discriminating items. In addition, over-achievers who selected as many as or less than the overlap point scored like under-achievers. The above groups were then labeled misclassified over-achievers ( $z_1$  and  $z_2$ ) and under-achievers ( $v_1$  and  $v_2$ ), respectively. The male and female misclassified over-achievers totaled forty-three and thirty-seven individuals, respectively. Male and female misclassified under-achievers totaled forty-four and sixty-eight persons, respectively. However, complete test data were not available for all subjects. Consequently, the sample was reduced to thirty-three male and thirty-two female over-achievers and forty-one male and forty-nine female under-achievers.

The sample reduction was added evidence of the misclassification of the discrepant achievement groups. In all other analyses sample was reduced more frequently because of incomplete testing information for the under-achiever than the over-achiever.

All of the groups were randomly divided into two equal samples for validation and cross-validation analyses purposes.

### F Item Overlap

An F item frequency distribution for all four inventories was constructed for the misclassified and properly classified male and female samples. The point of overlap where misclassified under-achievers scored on F items as properly classified under-achievers was determined by plotting the respective normal distribution curves.

The point of overlap between the groups was identified as three F items for the males and six F items for the females. Both overlap points were identical after replication of the procedure. The same procedure was followed using misclassified and properly classified over-achievers. The overlap point was identified as two F items for the males and four F items for the females which again held after cross-validation.

A t-test was used to determine significant differences between F item means of the male and female over-achieving misclassified and properly classified groups. A similar procedure was followed using under-achieving misclassified and properly classified samples. Comparisons of the significant findings with other pertinent data are given in Tables IX and X.

The means for the eight groups were in the theoretically predicted direction: both male and female under-achievers tend to select more F items than over-achievers. Properly classified under-achievers scored significantly higher on F items than the properly classified over-achievers. The misclassified over-achievers, who actually scored as under-achievers on GSCI items, scored significantly higher on F items than the properly classified over-achievers. The misclassified under-achiever, who actually scored as an over-achiever on GSCI items, selected significantly fewer F items than the properly classified

Table IX. Comparisons of F Item Means, Mean Squares and Sample Number Between Male and Female Misclassified and Properly Classified Groups

	Males	Females
<u>Misclassified</u>		
<u>Over-Achievers</u>	$\bar{X} = 2.61$ MS = 13.2121 N = 33	$\bar{X} = 6.45$ MS = 100.2812 N = 32
<u>Properly Classified</u>		
<u>Over-Achievers</u>	$\bar{X} = 1.03$ MS = 2.0606 N = 33	$\bar{X} = 2.25$ MS = 7.1875 N = 32
<u>Misclassified</u>		
<u>Under-Achievers</u>	$\bar{X} = 2.07$ MS = 9.2439 N = 41	$\bar{X} = 4.14$ MS = 35.1632 N = 49
<u>Properly Classified</u>		
<u>Under-Achievers</u>	$\bar{X} = 3.59$ MS = 22.0243 N = 41	$\bar{X} = 7.33$ MS = 96.4693 N = 49

under-achiever. However, misclassified under-achievers selected significantly more rarity items than the properly classified over-achievers.

Although there were no significant differences between means of the other groups, the direction of mean magnitude gives slight support to the underlying hypothesis. The misclassified over-achiever tended to select more rarity items than the misclassified under-achiever. In addition, there were no significant differences in F item selection between misclassified over-achievers and properly classified under-achievers.

The magnitude of differences of F item selection between the groups would have doubtless been greater if the population (from which the investigation sample had been drawn) had not been screened previously.



Table X. Comparisons of T-Values and Significant Levels of One Tailed Tests Between Male and Female Misclassified and Properly Classified Groups

		df	t-Value	Signifi- cance Levels (Percent)
<u>Males</u>				
Properly Classified Under-Achievers and	Properly Classified Over-Achievers	72	3.0117	.0005
Misclassified Under-Achievers and	Misclassified Over-Achievers	72	.6836	----
Misclassified Over-Achievers and	Properly Classified Over-Achievers	64	2.2898	.01
Misclassified Under-Achievers and	Properly Classified Under-Achievers	80	1.7272	.025
Misclassified Over-Achievers and	Properly Classified Under-Achievers	72	.9800	----
Misclassified Under-Achievers and	Properly Classified Over-Achievers	72	1.7931	.025
<u>Females</u>				
Properly Classified Under-Achievers and	Properly Classified Over-Achievers	79	2.8166	.005
Misclassified Under-Achievers and	Misclassified Over-Achievers	79	1.2762	.12
Misclassified Over-Achievers and	Properly Classified Over-Achievers	62	2.2581	.02
Misclassified Under-Achievers and	Properly-Classified Under-Achievers	96	1.9273	.03
Misclassified Over-Achievers and	Properly Classified Under-Achievers	79	.3832	----
Misclassified Under-Achievers and	Properly Classified Over-Achievers	79	1.6725	.05

Protocols of manifested uncooperative test-takers and individuals with erratic test performance were removed along with tests displaying obvious clerical errors before the present investigation was initiated. Hence, through earlier screening, the individuals whom the F scale attempts to identify were removed. However, regardless of the previous screening significant differences on F scale performance were obtained.

#### Effect of F on Test Reliability

It was hypothesized that the effectiveness of the F scale could be determined by its ability to remove unstable individuals who tend to lower instrument reliability by erratic test performance. Theoretically, reliability should increase with exclusion of unreliable subjects. However, the effect of homogeneity may operate also to reduce reliability. The question was asked which has the greater effect on reliability: erratic test performance or homogeneity? To test the effects of both of the above statements a random sample of subjects equal in magnitude to those identified by F as high fake potential were excluded. The assumption was made that the correlation reduced by random selection should be greater than the correlation reduced by F.

The critical F score obtained through the above procedure was then applied to a male and female stratified random sample of 66 over-achievers, under-achievers and normals. Sixteen individuals possessing total F scores as large as or greater than the critical F score were then excluded from the sample. In addition, an equal sized sample was randomly excluded to determine its effect on homogeneity. Hoyt's analysis of variance technique for estimating internal consistency reliability was used to obtain an estimate of consistency of the GSCI discriminating items before and after application of the F scale. The effects on reliability after application of the F scale are summarized in Table XI.

Table XI. Effects on GSCI Internal Consistency Reliability Before and After Application of the Male and Female F Scale

	Before Application of F		After Application of F	
	n	r	n	r
<u>Males</u>				
Total Sample	66	.85	50	.83
Randomly Excluded Sample	50	.85		
<u>Females</u>				
Total Sample	66	.78	44	.63
Randomly Excluded Sample	44	.69		

Although no significant differences between reliability coefficients were found, the magnitudinal direction of the coefficients did not substantiate the hypothesis: as previously stated, after application of the F scale instrument reliability should increase.

#### Effect of F on Test Validity

The effect of the F scale on the validity coefficient between GSCI raw scores and standardized grade point averages was also determined. It was hypothesized that the Pearsonian validity coefficient between the two variables would increase after application of the F scale. The effects of F on validity for males and females are summarized in Tables XII and XIII.

Table XII. Effects on the Validity Coefficient Between GSCI Raw Scores and Standardized Grade Point Averages After Application of The Male and Female F Scale

	n	r	Level of Significance p = 0 (Percent)
<u>Males</u>			
Correlation Before Application of F	66	.582	.005
Correlation After Application of F	50	.501	.005
Correlation After Random Exclusion	50	.564	.005
<u>Females</u>			
Correlation Before Application of F	66	.243	.03
Correlation After Application of F	44	.394	.005
Correlation After Random Exclusion	44	.322	.025

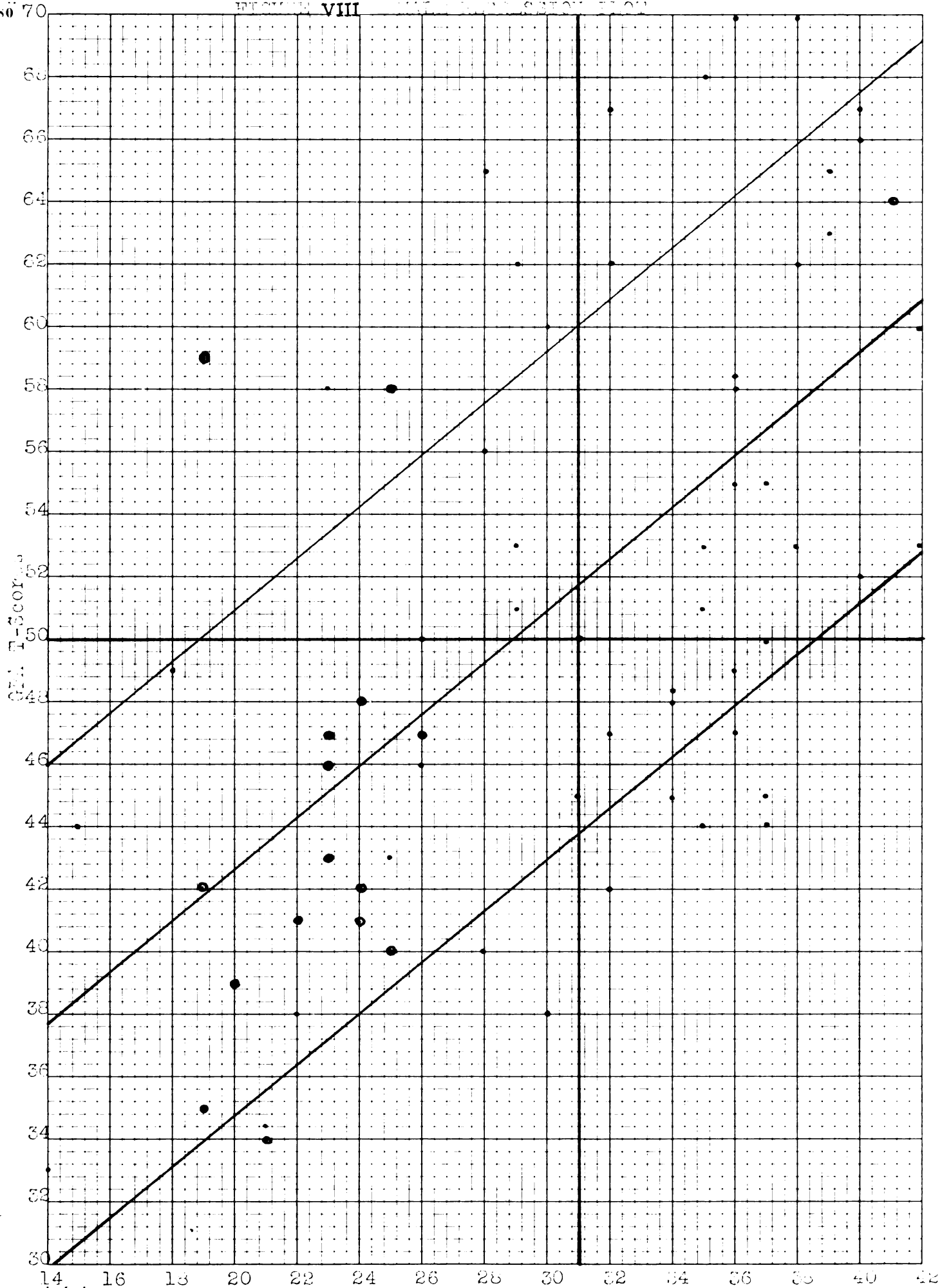
A linear regression line was plotted using GSCI raw scores and standardized grade point averages to locate placement of high F score males and females (see Figures VIII and IX). Eighteen percent of the males and thirty-eight percent of the females selecting a high number of rarity items fell one standard error of estimate below or above the regression line. Eighty-two percent of the high F males and forty-one percent of the high F females fell in the lower left quadrant of the regression plot. This area represents location of low achieving students.

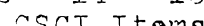
Table XIII. Significance of Difference Between Validity Correlation Coefficients Before and After Application of the Male and Female F Scale

	r After Application of F	r After Random Exclusion
<u>Males</u>		
Correlation Before Application of F	.59* (.27%)**	.11* (.45%)**
Correlation After Application of F		.45* (.32%)**
<u>Females</u>		
Correlation Before Application of F	.84*(.20%)**	.41* (.34%)**
Correlation After Application of F		.43* (.33%)**

\*  
z Transformation Scores

\*\*  
Level of Significance





### Summary

The outcomes of the present investigation were presented in this chapter. Selection of the items comprising the final F scale was based upon commonly selected rarity responses between validation and cross validation samples. In the final form the male F scale consists of twenty-five F items. The female F scale is composed of seventy-three F items.

Sex differences in F item selection were determined. Males selected five rarity items which were not chosen by females. Females selected fifty-three rarity items which were not chosen by males. Both males and females selected in common twenty F items.

Hoyt's analysis of variance technique was used in determining F scale reliability. For the males a reliability coefficient of .729 was obtained. For the females a reliability coefficient of .746 was determined. For scaling purposes this is slightly less than desirable.

The critical F score for both males and females was determined by plotting respective F distribution curves for misclassified and properly classified over- and under-achievers. The point of overlap where misclassified under-achievers scored as properly classified under-achievers on F items was identified as three rarity responses for males and six for females after cross-validation.

To obtain evidence of F scale validity three approaches were examined for the effect of F on: 1) expectancy of response fake; 2) test reliability; and 3) test validity.

Under-achievers selected significantly more F items than over-achievers in both male and female samples. Consequently, the rational of high fake expectancy was clearly substantiated.

The respective critical F scores were applied to a sample of males and females. Individuals possessing F scores as large as or greater than



the critical score were excluded from the sample. Hoyt's analysis of variance technique for estimating internal consistency was used to obtain a reliability statement of the GSCI discriminating items before and after application of the F scale. For both male and female samples no significant differences in reliability coefficients were obtained.

The effects on validity correlation of GSCI raw scores and standardized grade point averages before and after application of the male and female F scale were determined. Before application of the male and female validity coefficients were .582 and .243, respectively. After use of the F scale the male correlation decreased to .501 and the female validity coefficient increased to .394. However, no significant differences in correlations were obtained after F was applied. All validity correlations were significant from zero at the 3% or better level of confidence.

A linear regression was plotted using GSCI raw scores and standardized grade point averages to locate placement of high F score males and females. Eighteen percent of the males and thirty-eight percent of the females selecting a high number of F items fell one standard error of estimate below or above the regression line. Eighty-two percent of the high F males and forty-one percent of the high F females fell in the lower left quadrant of the regression plot. This area represents location of low achieving, low ability students.

The interpretation of the findings and the summary of the investigation are presented in Chapter V.

## CHAPTER V

### SUMMARY AND CONCLUSIONS

#### The Problem

This investigation was concerned with the development and validation of an F scale for an objective test battery of motivation.

#### Methodology

The present investigation used the administered test protocols of participants in Farquhar's motivational research project. Approximately 4200 eleventh grade Michigan public high school students comprised the population from which sample was drawn for this study.

Instrumentation consisted of the Generalized Situational Choice Inventory, the Preferred Job Characteristics Scale, the Human Trait Inventory and The Word Rating List. The battery consists of 502 items.

Rarity responses (based upon a 10% or less criterion for selection of the item) were determined separately for a stratified random sampling of 132 males and 132 females. A cross-validation sample of equal numbers was also used. Items comprising the final F scale were based upon commonly selected rarity responses between validation and cross validation samples. In the final form the male F scale consists of twenty-five F items. The female F scale is composed of seventy-three F items.

Sex differences in F item selection were determined. Males selected five rarity items which were not chosen by females. Females selected fifty-three rarity items which were not chosen by males. Both males and females selected in common twenty F items.

Hoyt's analysis of variance technique was used in estimating F scale reliability. For the males a reliability coefficient of .729 was obtained. For the females a reliability coefficient of .746 was determined.

The critical F score for both males and females was determined by plotting respective F distribution curves for misclassified and properly classified over- and under-achievers. The point of overlap where misclassified under-achievers scored as properly classified under-achievers on F items was identified as three rarity responses for males and six for females after cross validation.

To obtain evidence of F scale validity three approaches were examined for the effect of F on: 1) expectancy of response fake; 2) test reliability; and 3) test validity.

Under-achievers selected significantly more F items than over-achievers in both male and female samples. Consequently, the rational of high fake expectancy was clearly substantiated.

The respective critical F scores were applied to a sample of males and females. Individuals possessing F scores as large as or greater than the critical score were excluded from the sample. Hoyt's analysis of variance technique for estimating internal consistency was used to obtain a reliability statement of the GSCI discriminating items before and after application of the F scale. It was hypothesized that further evidence of the effectiveness of the F scale could be determined by its ability to remove unstable individuals who tend to lower instrument reliability by erratic test performance. Theoretically, reliability should increase with exclusion of unreliable subjects. However, the effect of homogeneity of test performance may operate also to reduce reliability. The question was raised as to which has the greater effect on reliability: erratic test performance or homogeneity of test performance. To test the effects of the

above question a random sample of subjects equal in magnitude to those identified by F as high fake potential were excluded. The assumption was made that the internal consistency reliability coefficient reduced by random selection should be greater than the reliability coefficient reduced by F selection.

Although no significant differences between reliability coefficients were found, the magnitudinal direction of the coefficients did not substantiate the hypothesis: as previously stated, after application of the F scale instrument reliability should increase.

The effects on validity between GSCI raw scores and standardized grade point averages before and after application of the male and female F scale were determined. Before application of F the male and female validity coefficients were .582 and .243, respectively. After use of the F the male correlation decreased to .501 and the female validity coefficient increased to .394. However, no significant differences in correlations were obtained after F was applied. All correlations were significant from zero at the 3% or better level of confidence.

A linear regression line was plotted using GSCI raw scores and standardized grade point averages to locate placement of high F score males and females. Eighteen percent of the males and thirty-eight percent of the females selecting a high number of rarity items fell one standard error of estimate below or above the regression line. Eighty-two percent of the high F males and forty-one percent of the high F females fell in the lower left quadrant of the regression plot. This area represents location of low achieving, low ability students.

### Conclusions

The following conclusions are based upon the findings of the investigation:

1. Females selected approximately three times as many rarity items as males. Males displayed a greater controversiality in response to test battery items. Conversely, females were significantly in greater agreement than males. The F scale therefore represents a measure of the presence of or lack of social conformity.

2. Males selected five rarity items which were not chosen by females. Females selected fifty-three rarity items which were not chosen by males. Both males and females selected in common twenty F items. The F scale therefore possesses the ability to tap an academic masculinity-femininity continuum.
3. The hypothesis that identified under-achievers would select more F items than identified over-achievers was significantly substantiated. The under-achiever selects significantly more F items because of the non-conformity, unstable characteristics of his behavior.
4. The F scale is able to identify significantly male and female properly classified under-achievers from properly classified over-achievers, misclassified over-achievers from properly classified over-achievers, misclassified under-achievers from properly classified under-achievers and misclassified under-achievers from properly classified over-achievers.
5. The hypothesis that GSCI reliability would increase after exclusion of high F score individuals was not substantiated. The assumption that correlation reduced by random selection should be greater than correlation reduced by F selection was not significantly supported.
6. A greater decrease of reliability coefficient would have doubtless occurred if the population (from which the investigation sample had been drawn) had not been screened previously. Protocols of manifested uncooperative test-takers and individuals with erratic test performance were removed along with tests displaying obvious clerical errors before the present investigation was initiated. Hence, through earlier screening, the very individuals whom the F scale attempts to identify were removed. However, regardless of the previous screening significant differences between under- and over-achievers on F scale performance were obtained.
7. The hypothesis that the correlation between GSCI raw scores and standardized grade point averages would increase after application of the F scale was not significantly substantiated. The male validity coefficient decreased while the female coefficient increase appreciably. Although no significant differences between validity coefficients occurred, the magnitudinal direction of the female coefficient supported the hypothesis. The lack of range of the male F scale (only 25 F items from the test battery) doubtless penalized the effectiveness of F to raise the male validity coefficient.

8. The reason for lack of increase in male validity coefficient was determined by plotting the regression line. Only eighteen percent of the high F males fell one standard error of estimate below and above the regression line. Thus, by excluding 82% of the stable individuals the validity coefficient decreased because of restriction. Conversely, the female validity coefficient increased because thirty-eight percent of the high F individuals fell one standard error below and above the regression line. Thus, a larger group of unstable individuals were excluded.
9. The male F scale successfully identified 60% of the individuals falling in the lower left quadrant of the regression plot. This area represents low achieving males. Consequently, the F scale appears to be an effective instrument in identification of low achieving males.
10. Further investigation with the F scale should be conducted before actual employment of the scale in test battery interpretation. Although there is significant empirical evidence to show that the F scale can differentiate reliably between under- and over-achievers, the complete validation of the instrument is lacking. For females the F appreciably but not significantly increases validity; for males the evidence is not so clear.
11. Re-evaluation of the F scale concept as used in the MMPI should be conducted. Lack of MMPI F scale validity casts serious doubt on its utility as a validation scale. In addition, from the evidence presented in this study scoring in a rarity direction does not necessarily preclude normalcy or valid protocols. Evidence supports the conclusion that rarity of response is not a basis for a validation key. F appears to possess potential for a discriminating scale between various subtle behavioral phenomena.

#### Implications for Further Research

Implications for further investigation include:

1. Increase male F item selection criterion to a fifteen percent level of response frequency. With increase magnitudinal range the male F scale would doubtless influence more decidedly test battery interpretation.
2. Analyze factorially the structure of the items comprising the F scale.

3. Determine the potential of the F scale as a test of conformity. Subjects with high agreement scores (as opposed to F item low agreement criterion) should be investigated to determine relationship between frequency items and conformity characteristics.
4. Examine behavioral characteristics of individuals selecting a high frequency of rarity items.
5. Construct and validate a research masculinity-femininity scale using male and female F scale items.
6. Examine item response and location of individual scores within the various quadrants of a regression plot to determine commonality of variables impinging upon behavioral characteristics.
7. Develop and validate as research instruments a response bias scale, a K scale (items on which under-achievers score as over-achievers, etc.), and a lie scale for the motivational battery.
8. Construct an adolescent independent-dependent striving response bias scale by using test items which have implications for tapping the above respective variables.
9. Construct an adolescent power-striving response bias scale by using the above procedure. Test items which possess implications for tapping adolescent power-needs would be determined and placed in a power scale.
10. Conduct rigorous validation investigations of validity scales on response distortion.

## BIBLIOGRAPHY



## BIBLIOGRAPHY

### A. BOOKS

- Edwards, A. L. Edwards Personal Preference Schedule. New York: The Psychological Corporation, 1959. 27 pp.
- Fricke, B. G. The Opinion, Attitude and Interest Survey. Minneapolis: Investors Diversified Services, 1955.
- Hartshorne, Hugh and May, M. A. Studies in Deceit. New York: Macmillan, 1928. 248 pp.
- Hathaway, S. R. Supplementary Manual for the MMPI. New York: The Psychological Corporation, 1946.
- Hathaway, S. R. and McKinley, J. C. The Minnesota Multiphasic Personality Schedule. Minneapolis: University of Minnesota Press, 1942.
- Hathaway, S. R. and McKinley, J. C. Manual for the MMPI. New York: The Psychological Corporation, 1946.
- Maller, M. B. Character Sketches. New York: Bureau of Publications, Teachers College, Columbia University, 1932. 388 pp.
- Maller, J. B. "Personality Tests." In J. M. Hunt, Personality and the Behavior Disorders. New York: Ronald Press, 1944.
- Kuder, G. F. Kuder Preference Record Vocational. Chicago: Science Research Associates, 1956. 35 pp.
- Kuder, G. F. Kuder Preference Record Occupational. Chicago: Science Research Associates, 1959. 18 pp.
- Ruch, F. L. "A Technique for Detecting Attempts to Fake Performance on a Self-Inventory Type of Personality Test." In Quinn McNemar and M. A. Merrill, Studies in Personality. New York: McGraw-Hill, 1942.
- Strong, E. K. Vocational Interest of Men and Women. Stanford: Stanford University Press, 1943. 746 pp.
- Symonds, P. M. Diagnosing Personality and Conduct. New York: Appleton-Century, 1932. 602 pp.

## B. PERIODICALS

- Adams, C. R. "A New Measure of Personality," Journal of Applied Psychology, 1941, 25:141-151.
- Allport, G. W. "A Test for Ascendance-Submission," Journal of Abnormal Psychology, 1928, 23:118-136.
- Allport, G. W. "The Use of Personal Documents in Psychological Science," Social Science Research Council Bulletin, 1942, Number 42.
- Benton, A. L. "The Interpretation of Questionnaire Items in a Personality Inventory," Archives of Psychology, 1935, Number 190.
- Benton, A. L. "The MMPI in Clinical Practice," Journal of Nervous and Mental Disorders, 1945, 102:416-420.
- Bernreuter, R. G. "Validity of the Personality Inventory," Personality Journal, 1933, 11:383-386.
- Bernreuter, R. G. "Theory and Construction of the Personality Inventory," Journal of Social Psychology, 1933, 4:387-405.
- Bernreuter, R. G. "The Present Status of Personality Trait Tests," Educational Research Supplement, 1940, 21:160-171.
- Bills, Marion. "Selection of Casualty and Life Insurance Agents," Journal of Applied Psychology, 1941, 25:6-10.
- Bordin, E. S. "A Theory of Vocational Interests as Dynamic Phenomena," Educational and Psychological Measurement, 1943, 3:49-65.
- Cady, V. M. "The Estimation of Juvenile Incurability," Journal of Delinquency Monographs, 1923, Number 2.
- Cofer, C. N., Chance, J. and Judson, A. J. "A Study of Malingering on the MMPI," Journal of Psychology, 1949, 27:491-499.
- Cottle, W. C. "Card Versus Booklet Forms of the MMPI," Journal of Applied Psychology, 1950, 34:255-259.
- Cottle, W. C. "The MMPI: A Review," Kansas Studies in Education, 1953, 3:1-82.

- Cronbach, L. J. "Response Sets and Test Validity," Educational and Psychological Measurement, 1946, 6:475-494.
- Cronbach, L. J. "Further Evidence on Response Sets and Tests Designs," Educational and Psychological Measurement, 1950, 10:3-31.
- Eisenberg, P. "Individual Interpretation of Psychoneurotic Inventory Items," Journal of Genetic Psychology, 1941, 25:19-40.
- Eisenberg, P. and Wesman, A. "A Consistency in Responses and Logical Interpretation of Psychoneurotic Inventory Items," Journal of Educational Psychology, 1941, 32:321-338.
- Frenkel-Brunswik, E. "Mechanisms of Self-Deception," Journal of Social Psychology, 1939, 10:409-420.
- Fricke, B. G. "Conversion Hysterics and the MMPI," Journal of Clinical Psychology, 1956, 12:322-326.
- Fricke, B. G. "A Response Bias Scale for the MMPI," Journal of Counseling Psychology, 1957, 4:149-153.
- Fricke, B. G. "Subtle and Obvious Test Items and Response Set," Journal of Consulting Psychology, 1957, 21:250-252.
- Gough, H. G. "Simulated Patterns on the MMPI," Journal of Abnormal and Social Psychology, 1947, 42:215.
- Gough, H. G. "Factors Relating to the Academic Achievement of High School Students," Journal of Educational Psychology, 1949, 40:65-78.
- Gough, H. G. "What Determines the Academic Achievement of High School Students," Journal of Educational Research, 1953, 46:321-331.
- Gough, H. G. "The F Minus K Dissimulation Index for the MMPI," Journal of Consulting Psychology, 1950, 14:408-413.
- Guilford, J. P. and Guilford, R. B. "Personality Factors S, E, and M and Their Measurement," Journal of Psychology, 1936, 2:109-127.
- Hathaway, S. R. and McKinley, J. C. "A Multiphasic Personality Schedule: I. Construction of the Schedule," Journal of Psychology, 1940, 10:249-254.

- Hathaway, S. R. and McKinley, J. C. "A Multiphasic Personality Schedule: III. The Measurement of Symptomatic Depression," Journal of Psychology, 1942, 14:73-84.
- Horst, Paul. "The Prediction of Personal Adjustment," Social Science Research Council Bulletin, 1941, Number 48.
- Hovey, H. B. "Detection of Circumvention in the MMPI," Journal of Clinical Psychology, 1948, 4:97.
- Hoyt, C. J. "Test Reliability Estimated by Analysis of Variance," Psychometrika, 1941, 6:133-160.
- Humm, D. G. and Humm, K. A. "Validity of the Humm-Wadsworth Temperament Scale: With Consideration of the Effects of Subjects' Response-Bias," Journal of Psychology, 1944, 18:55-64.
- Humm, D. G., Stormont, R. C. and Iorns, M. E. "Combination Scores for the Humm-Wadsworth Temperament Scale," Journal of Psychology, 1939, 7:227-253.
- Humm, D. G. and Wadsworth, G. W. "The Humm-Wadsworth Temperament Scale," American Journal of Psychiatry, 1935, 92:163-200.
- Hunt, H. F. "A Study of the Differential Diagnostic Efficiency of the MMPI," Journal of Consulting Psychology, 1948, 12:331-336.
- Hunt, H. F. "The Effect of Deliberate Deception on MMPI Performance," Journal of Consulting Psychology, 1948, 12:396-402.
- Hunt, W. A. "The Detection of Malingering: A Further Study," United States Naval Medical Bulletin, 1946, 46:249.
- Hunt, W. A. and Older, H. J. "Detection of Malingering Through Psychometric Tests," United States Naval Medical Bulletin, 1943, 41:1318.
- Kazan, A. T. and Sheinberg, I. M. "Clinical Note on the Significance of the Validity Score F in the MMPI," American Journal of Psychiatry, 1945, 102:181-183.
- Kelly, E. L., Miles, C. C. and Terman, L. M. "Ability to Influence One's Score on a Typical Pencil and Paper Test of Personality," Character and Personality, 1936, 4:206-215.

- Krumboltz, J. D. and Farquhar, W. W. "The Effect of Three Teaching Methods on Achievement and Motivational Outcomes in a How-to-Study Course," Psychological Monographs, 1957, 71: Number 14.
- Laird, D. A. "Detecting Abnormal Behavior," Journal of Abnormal Psychology, 1926, 20:128-141.
- Landis, C. and Katz, S. E. "The Validity of Certain Questions Which Purport to Measure Neurotic Tendencies," Journal of Applied Psychology, 1934, 18:343-356.
- Levine, A. S. "A Technique for Developing Suppression Tests," Educational and Psychological Measurement, 1952, 12:313-315.
- Maller, J. B. "The Effect of Signing One's Name," School and Society, 1930, 31:882-884.
- McKinley, J. C. and Hathaway, S. R. "A Multiphasic Personality Schedule: V. Hysteria, Hypomania and Psychopathic Deviate," Journal of Applied Psychology, 1942, 14:73-84.
- McKinley, J. C., Hathaway, S. R. and Meehl, P. E. "The MMPI: VI. The K Scale," Journal of Consulting Psychology, 1948, 12:20-31.
- McNemar, Quinn. "The Mode of Operation of Suppressant Variables," American Journal of Psychology, 1945, 58:554-555.
- McQuary, J. J. and Truax, W. E. "An Under Achievement Scale," Journal of Educational Research, 1955, 48:393-399.
- Meehl, P. E. "An Investigation of General Normality Control Factor in Personality Testing," Psychological Monographs, 1945, 59: Number 4.
- Meehl, P. E. "The Dynamics of Structured Personality Tests," Journal of Clinical Psychology, 1945, 1:296-303.
- Meehl, P. E. and Hathaway, S. R. "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:525-564.
- Metfessel, M. "Personality Factors in Motion Picture Writing," Journal of Social and Abnormal Psychology, 1935, 30:333-347.

- Middleton, George and Guthrie, G. M. "Personality Syndromes and Academic Achievement," Journal of Educational Psychology, 1959, 50: Number 2.
- Mosier, C. I. "A Note on Item Analysis and the Criterion of Internal Consistency," Psychometrika, 1936, 1:275-282.
- Olson, W. C. "The Waiver of Signature in Personal Reports," Journal of Applied Psychology, 1936, 20:442-450.
- Ossipov, V. P. "Malingering: The Simulation of Psychosis," Bulletin of the Menninger Clinic, 1944, 8:39-42.
- Rosen, E. "Self Appraisal and Perceived Desirability of MMPI Personality Traits," Journal of Counseling Psychology, 1956, 3:44-51.
- Rosen, E. "Self Appraisal, Personal Desirability, and Perceived Social Desirability of Personality Traits," Journal of Abnormal and Social Psychology, 1956, 52:151-158.
- Rosenzweig, Saul. "A Suggestion for Making Verbal Personality Tests More Valid," Psychological Review, 1934, 41:400-401.
- Rosenzweig, Saul. "A Basis for the Improvement of Personality Tests with Special Reference to the M-F Battery," Journal of Abnormal and Social Psychology, 1938, 33:476-488.
- Schmidt, H. O. "Test Profiles as a Diagnostic Aid: The MMPI," Journal of Applied Psychology, 1945, 29:115-131.
- Schmidt, H. O. "Notes on the MMPI: The K Factor," Journal of Consulting Psychology, 1948, 12:337-342.
- Schneck, J. M. "Clinical Evaluation of the F Scale on the MMPI," American Journal of Psychiatry, 1948, 104:440-442.
- Shoben, E. J. "The Assessment of Parental Attitudes in Relation to Child Adjustment," Genetic Psychology Monographs, 1949, 39:101-148.
- Spencer, D. "Frankness of Subjects on Personality Measures," Journal of Educational Psychology, 1938, 29:26-35.
- Steinmetz, H. C. "Measuring Ability to Fake Occupation Interest," Journal of Applied Psychology, 1932, 16:123-230.

- Sweetland, A. "Hypnotic Neurosis-Hypochondriasis and Depression," Journal of Genetic Psychology, 1948, 39:19-105.
- Vernon, P. E. "The Attitude of the Subject in Personality Testing," Journal of Applied Psychology, 1934, 18:165-177.
- Washburne, J. N. "A Test of Social Adjustment," Journal of Applied Psychology, 1935, 19:125-244.
- Wherry, R. J. "Test Selection and Suppressor Variables," Psychometrika, 1946, 11:239-247.
- Wiener, D. N. "Selecting Salesmen with Subtle-Obvious Keys for the MMPI," American Psychologist, 1948, 3:364.
- Wiener, D. N. "Subtle and Obvious Keys for the MMPI," Journal of Consulting Psychology, 1948, 12:164-170.
- Wiener, D. N. "A Control Factor in Social Adjustment," Journal of Abnormal and Social Psychology, 1951, 46:3-8.
- Willoughby, R. R. and Morse, M. E. "Spontaneous Reactions to a Personality Inventory," American Journal of Orthopsychiatry, 1936, 6:562-575.

### C. UNPUBLISHED MATERIALS

- Arnold, D. A. "The Clinical Validity of the Humm-Wadsworth Temperament Scale in Psychiatric Diagnosis." Unpublished Doctor's dissertation, University of Minnesota, Minneapolis, 1942.
- Farquhar, William W. "A Comprehensive Study of the Motivational Factors Underlying Achievement of Eleventh Grade High School Students." East Lansing: Approved Research Application to the Commissioner of Education, United States Office of Education, 1959. (Mimeographed.)
- Fricke, B. G. "The Development of an Empirically Validated Personality Test Employing Configural Analysis for the Prediction of Academic Achievement." Unpublished Doctor's dissertation, University of Minnesota, 1954.

Hendrickson, G. "Attitudes and Interests of Teachers and Prospective Teachers." Paper read before Section Q, AAAS, Atlantic City, December 27, 1932.

Jeffery, M. E. "Some Factors Influencing Answers on the Multiphasic K Scale." Unpublished Doctor's dissertation, University of Minnesota, Minneapolis, 1946.



## APPENDIX

MOTIVATIONAL TEST BATTERY  
F SCALE ITEMS

0 = Male Rarity Item  
X = Female Rarity Item  
\* = Male and Female  
Rarity Item

Human Trait Inventory

Agree Direction

- 0 - 2 I like collecting flowers or growing house plants.  
\* - 25 I have played that I am sick to get out of doing something.  
\* - 52 Most of my school subjects are a complete waste of time.  
\* - 67 When I was a youngster I stole things.  
X - 79 I have played hooky from school.  
0 - 81 There was a time in my life when I liked to play with dolls.  
X - 85 My parents object to the friends I choose.  
\* - 87 I have been sent to the principal for misbehaving in class.  
\* - 88 I have trouble with my muscles twitching or jumping.  
X - 97 One or more times a week I suddenly feel hot all over  
for no apparent reason.  
X - 112 I wish I were a child again.  
0 - 115 I feel cross and grouchy without good reason.  
X - 120 I feel that I haven't any goals or purpose in life.  
X - 124 I would like to belong to a motorcycle club.

Word Rating List

Agree Direction

Teachers feel that I am:

- \* - 3 dull  
X - 4 inefficient  
\* - 11 unsuccessful  
\* - 14 "blah"

- X - 19    uninterested
- \* - 21    unreliable
- X - 37    childish
- X - 43    cold
- X - 44    below average
- \* - 45    reckless
- X - 51    a goof off
- X - 53    lazy
- X - 62    unreasonable
- X - 64    a "wheel"
- X - 65    a "grind"
- X - 66    fool-hearty
- X - 69    retiring
- X - 79    a "brain"
- \* - 90    outsider
- X - 92    a person who delays
- X - 93    indecisive
- \* - 94    irresponsible
- X - 108    fault-finding
- X - 110    dominant
- \* - 111    inaccurate
- X - 114    pushed

### Generalized Situational Choice Inventory

#### Agree Direction

I would prefer to:

- X - 2    Do well in school.
- X - 9    Be quick, but often incorrect.
- X - 62    Receive grades which are like everyone else's.
- X - 107    Be known for what I could do.
- \* - 114    Learn by defeating an inexperienced player.
- \* - 137    Have decisions made for me.

- X - 144 Live a life of leisure.
- X - 176 Have an instructor who gave me an "A" and not care whether I learned anything or not.
- X - 180 Accept what someone else says even though I don't agree.
- X - 187 Date the smartest girl or boy in class.

#### Disagree Direction

- 0 - 17 Make something planned by somebody else.
- \* - 20 Accomplish a task in a hurry, but less carefully.
- X - 50 Finish a job.
- 0 - 51 Play a game against inexperienced players and win.
- X - 77 Do a recognized but incomplete job.
- X - 88 Take a course from an instructor who only gives "C's".
- \* - 102 Have an easier job which pays less.
- X - 118 Buy something on credit and pay for it as I use it.
- X - 119 Do what others think is right.
- X - 122 Be known as an expert.
- X - 139 Accomplish a difficult task fast.
- X - 157 Win a game from an inexperienced player.
- X - 170 Have lots of money.

#### Preferred Job Characteristics Inventory

##### I Prefer:

- X - 1b A job with short working hours.
- \* - 7a A job which requires little thinking.
- \* - 8b A job where I make few if any decisions.
- X - 14b A job which requires little thinking.
- \* - 15b A job where I make few if any decisions.
- X - 17b A job where I could not be fired.

- X - 25b A job which permits me to take days off when I want.
- X - 39a A job where I could not be fired.
- X - 46a A job where I could not be fired.
- X - 49a A job which requires little thinking.
- X - 57b A job where I make few if any decisions.
- X - 64b A job which requires little thinking.

ROOM USE ONLY

~~JUN 21 1963~~

~~MAY 10 1964~~

~~AUG 10 1964~~

~~AUG 17 1964~~

~~SEP 1 1964~~

~~ALL 253~~ <sup>14</sup>  
90

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03177 5491